DOCTOR OF PHILOSOPHY

Assessment of potential antibacterial drug targets

Diaz Saez, Laura

Award date:
2016

A thesis submitted for the degree of Doctor of Philosophy

Assessment of Potential Antibacterial Drug Targets

Laura Díaz Sáez

Supervisor:

Professor William N. Hunter

School of Life Sciences

University of Dundee

November 2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF EQUATIONS AND TABLES

Equations

Tables

LIST OF ABBREVIATIONS

Abbreviations were conducted following the International Union of Pure and Applied Chemistry (IUPAC, www.chem.qmul.ac.uk/iupac) and the International Union of Biochemistry and Molecular Biology (IUBMB, www.chem.qmul.ac.uk/iubmb) nomenclature. Proteins and compound abbreviations follow BRENDA (www.brenda-enzymes.org) and Acronyms-Abbreviations (www.chemie.fu-berlin.de/cgi-bin/acronym) databases. Amino acids are represented with the three or one letter identifier, and atoms as their elemental symbol. Standard SI abbreviations are used except Ångstrom (Å, 10^{-10} m). Additional abbreviations as follows:

AAA+	ATPases associated with various cellular activities
AChRs	Acetylcholine receptors
ACP	Acyl-carrier-protein
Ba	*Bacillus anthracis*
Bc	*Burkholderia cenocepacia*
Bce	*Bacillus cereus*
BLI	Biolayer interferometry
Bm	*Bombyx mori*
Bp	*Burkholderia pseudomallei*
Bs	*Bacillus subtilis*
Bt	*Bos taurus*
D-Ala:LA	D-Ala bound to LA by an ester bond
Cb	*Coxiella burneii*
CDC	The U.S. centers for disease control and prevention

DCS	D-cycloserine
DDM	*n*-dodecyl β-D-maltoside
DDU	Drug discovery unit
DEG	Database of essential genes
DHF	Dihydrofolate
DHPPP	6-hydroxymethyl-7,8-dihydropterin pyrophosphate
DLID	Drug-like density score
Dm	*Drosophila melanogaster*
DSF	Differential scanning fluorimetry
Dstl	Defence science and technology laboratory
Ec	*Escherichia coli*
ELF	European lead factory
FASI	Fatty acid biosynthetic pathway type I
FASII	Fatty acid biosynthetic pathway type II
FFT	Fast Fourier transform
gDNA	Genomic DNA
GSK	GlaxoSmithKline
Hp	*Helicobacter pylori*
H_2Pte	Dihydropteroate
Hs	*Homo sapiens*
HTP	High-throughput
InF	Intensity of fluorescence
LA	Lipoteichoic acid
Lc	*Lactobacillus casei*
Lm	*Listera monocytogenes*

MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight.
MATE	Multidrug and toxic compounds extrusion
mDAP	*meso*-diaminopimelate
MIC	Minimum inhibitory concentration
Ml	*Micrococcus luteus*
MRSA	Methicillin-resistant *Sa*
MS	Mass spectrometry
Mt	*Mycobacterium tuberculosis*
MTX	Methotrexate
N	Number of molecules in the asymmetric unit
NAcGlc	*N*-acetyl-glucosamine
NAcMur	*N*-acetyl-muramic acid
NCS	Non-crystallographic symmetry
NFK	*N*-formyl-L-kynurenine
Ng	*Neisseria gonorrhoeae*
Ni-NTA	Tris-nitriloacetic acid (tris-NTA) charged with Ni^{2+}
NMDARs	*N*-methyl-D-aspartate glutamate receptors
Pa	*Pseudomonas aeruginosa*
PDB	Protein data bank
PG	Peptidoglycan
Pf	*Plasmodium falciparum*
QED	Quantitative estimate of drug-likeness
Rm	*Ralstonia metallidurans*
r.m.s.d.	Root mean square deviation
Rn	*Rattus norveqieus*

SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
rRNA	Ribosomal RNA
Sa	*Staphylococcus aureus*
Sc	*Saccharomyces cerevisiae*
s.d.	Standard deviation
Se	*Salmonella enterica*
SEC	Size-exclusion chromatography
Sf	*Streptococcus faecalis*
Sm	*Streptococcus mutants*
Sp	*Streptomyces parvulus*
Sp	*Streptococcus pneumoniae*
TEV	Tobacco etch virus
THF	Tetrahydrofolate
TKO	Genome-transposon knock-outs
TraDIS	Transposon-directed insertion site sequencing
$\mathbf{V_m}$	Matthews coefficient
XANES	X-ray absorption near-edge structure
Xo	*Xantomonas oryzae*
Yp	*Yersinia pestis*

ACKNOWLEDGMENTS

I would like to express all my gratitude to the people without whom this research would not have been possible:

First to Bill Hunter who welcomed me in his lab and whom I thank for all the mentorship, advice, motivation, teaching and patience during these three years. Also for recommendations of unforgettable Scottish wonders.

To my labmates who always helped and taught me, and became valuable friends. My special thanks to Alice Dawson, who proof-read this thesis, and Paul Fyfe for their advice on X-ray diffraction data collection and structure determination. Also to Velupillai Srikannathasan for introducing me to the KynB project, Martin Zoltner for his teaching on membrane protein production and crystallisation, Thomas Eadsforth for advise on my first assay optimisation, and the Brazilian gang (Carla Gottschald, Gustavo Mercaldi and Renata Reis) for our science and life talks.

To Stuart McElroy and his team at the ELF, and the DDU screening team, in particular Leah Torrie, who gave advice and support on HTP assays and compound library screens. Also, special thanks to the Dstl for funding and their strategic contributions.

To Laura Riolobos (my first mentor), Mark van Raaij and Carmela García for showing me the fun of science when I was an undergraduate.

Last but not least, I would like to thank my parents (Charo and Luis), sister (Julia) and friends, especially to Jana Alonso, for all the unconditional support received during these years in both thesis and personal life. Thank you!

I will never forget your friendship and everything I have learnt. Thank you all for making these three years one of the greatest experiences I ever had.

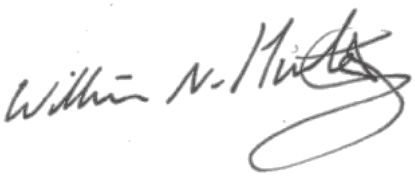
DECLARATION

I thereby declare that the following thesis is based on the results of investigations conducted by myself, and that this thesis is of my own composition. This thesis has not, in whole or part, been previously presented for a higher degree. Work other than my own is clearly indicated in the text by reference to the relevant researchers or publications. I would like to thank those people who allowed me to refer to their unpublished observations and data, which are individually acknowledged where they occur.

Laura Díaz Sáez

The work presented in this thesis is the work of the candidate Laura Díaz Sáez.

Conditions of the relevant Ordenance and Regulations have been fulfilled.



Professor William N. Hunter

ABSTRACT

The alarming increase in antibacterial drug resistance indicates an urgent need to develop new drugs. This project aimed to assess and select potential antibacterial targets and carry out initial biochemical characterisation concentrating on enzymes from biowarfare agents *Bacillus anthracis*, *Burkholderia pseudomallei*, *Francisella tularensis and Yersinia pestis*. The overall objective is to combine genetic and chemical studies to validate, or not, targets for early stage drug discovery.

In collaboration with the Dstl (Defence Science and Technology Laboratory), a series of potential targets were selected. This was carried out using essentiality prediction data from the Dstl labs, information at AEROPATH and ChEMBL databases, and the literature. Once ten different targets were selected, recombinant protein production was carried out to support structural and biochemical characterisation. Seven proteins were successfully purified and four of them prioritised for further studies. These are kynurenine formamidase (KynB), D-alanine—D-alanine ligase (Ddl), caseinolytic protease subunit P (ClpP), and the dihydrofolate synthase:folyl-poly-glutamate synthase (bifunctional protein FolC). X-ray crystallography was used to determine protein structures of KynB, Ddl and ClpP from various bacteria. Additionally, different enzymatic and binding assays were applied to assess kinetic parameters of KynB, Ddl and FolC, and compound library screenings were carried out for Ddl. In parallel, the genetic validation of these targets was being carried out by the Dstl.

KynB is an important enzyme in tryptophan metabolism and predicted to be essential in *Pseudomonas aeruginosa*. The *B. anthracis*, *B. cenocepacia* and *P. aeruginosa* KynB structures showed an amidase fold not previously described, with a distinctive binuclear

Zn^{2+} catalytic site that indicated a distinct reaction mechanism. Whilst the characterisation of the enzyme was ongoing, the Dstl lab reported the gene as non-essential and so no additional chemical validation was pursued. Ddl generates a precursor for the peptidoglycan layer and appears to be an essential protein in several Gram-negative bacteria. The structure of *B. pseudomallei* Ddl (*Bp*Ddl) in the presence of the co-factor and the reaction product D-alanyl-D-alanine (1.5 Å resolution) gives information about the substrate-binding site. Biolayer interferometry (BLI) and high-throughput (HTP) assay protocols were developed and applied. Despite testing around 22,000 compounds no inhibitors or suitable hits were found. This suggests *Bp*Ddl is a challenging target and a different approach for drug discovery might have to be considered. Fourteen ClpP subunits form a proteolytic complex which presents two different conformations; open and compressed. Structures of *F. turalensis* ClpP (*Ft*ClpP) with open and compressed conformations have been determined indicating this major conformational change is caused by a loop rearrangement at the proteasome inner canal that leads to the protease active site. FolC is an essential protein for the synthesis of folyl-poly-glutamates, a reference pathway for drug development. A new enzymatic assay, using malachite green, has been identified and used to confirm *Y. pestis* FolC activity. Such an assay could be used to determine kinetic parameters and to develop a HTP assay.

The studies carried out informed about potential antibacterial target structures and biochemical properties. Protocols have been developed for protein recombinant expression, purification, crystallisation and structure determination as well as enzymatic assay development and compound library screens.

PART ONE

INTRODUCTION

1.1 Antibacterial drugs and resistance

During the last century important discoveries have been made in the development of antibacterial drugs and different classes of antibiotics have been successfully used (Fair and Tor, 2014). These include aminoglycosides (*e.g.* streptomycin, gentamycin and kanamycin), amphenicols (*e.g.* chloramphenicol), β-lactams (*e.g.* penicillin G, ampicillin and amoxicillin), glycopeptides (*e.g.* vancomycin and telavancin), lipopeptides (*e.g.* daptomycin), macrolides (*e.g.* erythromycin and azithromycin), quinolones (*e.g.* ciprofloxacin and levoflaxacin), rifamycins (*e.g.* rifampicin), streptogramins (*e.g.* dalfopristin), sulphonamides (*e.g.* trimethoprim) and tetracyclines (*e.g.* tetracycline and doxycycline). However, the alarming increase of (multi)drug resistance to these commonly used treatments is a major global healthcare problem (Silver, 2011; Chopra, 2013; Fair and Tor, 2014). The natural occurrence of drug resistance mechanisms plus the inappropriate use of antibiotics, for example by over-prescription, patients failing to complete treatments or use in agriculture growth promotion, are the main causes of the increase in antibacterial resistance (Chopra, 2013; Done *et al*., 2015). This is reflected in increasing clinical isolates presenting drug resistance (Brooks and Brooks, 2014; Matthew *et al.,* 2015).

A wide range of drug resistance mechanisms, constitutively present or induced under stress conditions, have been described and the set of components involved is called the resistome (Silver, 2011). These mechanisms include chemical modification of the drug or sequestration (a non-essential protein binds to the drug), mutation of the targeted

protein or increasing its expression level, modification of the enzyme (*e.g.* methylation) to hinder drug binding, reduction of drug uptake (*e.g.* down regulation of porin expression), and increasing the efflux (*e.g.* overexpression of efflux pumps) (Li *et al.,* 2005; Li and Nikaido, 2009; Silver *et al.,* 2012; Torok *et al.,* 2012). Specifically in Gram-positive bacteria, modification of the properties of the peptidoglycan layer triggers a change in susceptibility to drugs. Resistance mechanisms can be transferred among bacteria by horizontal gene transmission, transformation (exogenous DNA), transduction (via bacteriophages) or conjugation (Culyba *et al.,* 2015). In some cases, a combination of such mechanisms results in multidrug resistance (Li *et al.,* 2005). In order to address the complex issue of drug resistance it is important to find new antibacterial compounds.

1.2 Strategies for antibacterial drug development

1.2.1 The past

An early approach in antibiotic discovery was based on screens testing a collection of dyes and other chemicals. Paul Ehrlich defined the term chemotherapy as the use of synthetically produced chemicals to destroy pathogenic microorganisms, and worked on the development of selective biologically active compounds harmless for the host (Gelpi *et al.,* 2015). His studies resulted in the discovery of salvarsan (in 1907, Bosh and Rosich, 2008), an antibiotic against syphilis (Reithmiller, 2005; Gelpi *et al.,* 2015). Following the discovery of penicillin (Fleming, 1929), natural products from fermentation broths and extracts of microorganisms were introduced in screening. This became the start of the "Golden Age" in drug discovery (from about 1940 to 1970), characterised by the description of a high number of new antibacterial classes (Silver, 2012; Figure 1.1). In the early 1960s, compound libraries were modified to detect

pathway-specific and intensive screens searching for inhibitors targeting the cell wall biosynthetic pathway produced three new classes of antibiotics; monobactams, carbapenems and fosfomycin (Silver, 2011). The single-component directed screening approach was proposed in 1977 (Cohen, 1997) and turned drug discovery into what it is implemented nowadays (Silver, 2011).

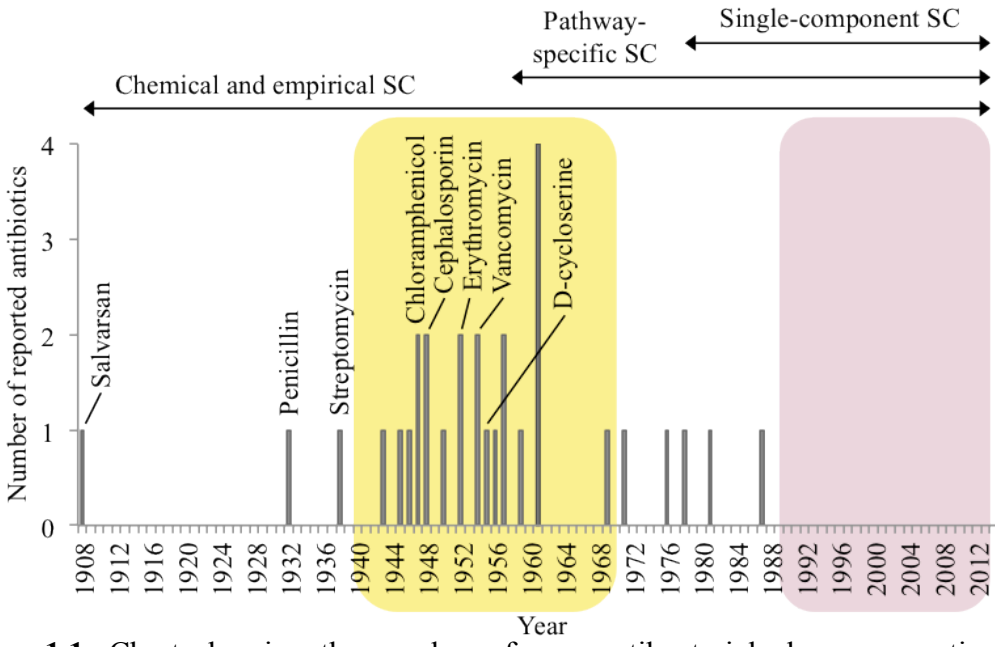


Figure 1.1: Chart showing the number of new antibacterial classes over time and different strategies used for drug search. SC is the abbreviation for screening. The Golden Age is coloured in yellow and the Void Age in purple. Information from 1908 until 2012 has been obtained from Adamczyk-Woźniak *et al.,* 2009; Silver, 2012.

1.2.2 The present

1.2.2.1 The Void Age

Despite improvements in technologies associated with drug discovery, the accessibility of protein structures and large-scale genomic and chemical data, no new drug class has been developed over the past 25 years, the so-called "Void Age" (Silver, 2011; Chopra, 2013). The lack of new drug classes has been reviewed and attributed mainly to the lack of investment from the pharmaceutical industry and funding bodies, decisions likely influenced by high cost, timescales and high risk (Gwynn *et al.,* 2010; Silver, 2011;

Chopra, 2013). The emerging antibacterial drug resistance crisis highlights the need for new drugs and to exit the Void Age.

Initiatives are being implemented seeking to increase social awareness, introducing new procedures to avoid the misuse of antibiotics (ReAct initiative by Europe by the British Society for Antimicrobial Chemotherapy, www.bsac.org.uk) and improving the current technologies to find 10 new antibacterial drugs by 2020 (10x20 initiative by the Infectious Diseases Society of America, www.idsociety.org). European organisations are working together and have created the Innovative Medicines Initiative (IMI, www.imi.europa.eu), a collaborative partnership between private and public organisations to focus efforts investigating the mechanisms of drug transport and to apply the findings to improve antibacterial drug development (Tommasi *et al.,* 2015).

Having recognised deficiencies in antibiotic discovery (see also section 1.2.2.2), the situation has improved with four new classes of compounds identified recently. These include aryl isonitrile (Davis *et al.,* 2015), ramizol (Iscla *et al.,* 2015), lysocin E (Hamamoto *et al.,* 2015) and teixobactin (Ling *et al*., 2015). A small library of aryl isonitrile derivatives was used in a phenotypic screen and this helped identify a potential antibiotic for skin infections of MRSA (Methicillin-resistant *Staphylococcus aureus*). Ramizol (1,3,5-tris[(1E)-2'-(4''-benzoic acid)vinyl]benzene) is the result of an *in silico* compound design process targeting the ion channel MscL of MRSA, and lysicin E has been obtained from an empirical screen using bacterial supernatants and seems to bind menaquinone (vitamin K), a component of the electron transport system. Texiobactin, a cell wall biosynthesis inhibitor, was also obtained from a phenotypic screen. In this case, soil bacteria were isolated and grown using a multichannel device, called iChip, to

find new antibiotic compounds from growth extracts. These new classes are in the early stage of development.

1.2.2.2 Drug discovery strategies

Different strategies are used to search for new antibacterial drugs, and these can be classified depending on the screening method (phenotypic or single-target), and the nature of the chemical library (diverse or target-based synthetic compounds, and natural products) (Copeland, 2005). Drug development can be focused on the determination of new active compounds for known targets (a protein which has its activity modified by the drug, that hinders cell growth or promotes cell death) or the search for compounds active against new antibacterial targets. Drug discovery targeting new systems is more likely to yield novelty in terms of chemical structure and mechanism of action. The benefits of such a development could be the absence of resistance mechanisms in the clinically relevant bacteria (Silver, 2011).

To select new potential targets, the most common approach is to consider gene essentiality experiments (further discussed in section 1.2.2.3). The potential targets then need to be validated employing two approaches: genetic and chemical (Frearson *et al.,* 2007). Genetic validation confirms the essentiality of the target for bacterial growth and/or survival using methodologies such as knock-outs and deleterious mutation generation (Freiberg and Brötz-Oesterhelt, 2005). Chemical validation involves the use of a compound capable of binding with high affinity to the target and producing concentration dependent cell death or inhibition of growth (Wyatt *et al.,* 2011).

Once the target has been selected, compounds that interact have to be found. High-throughput (HTP) assays are being developed to screen libraries to obtain leads, which

are a starting point for generating target-specific (or single-target) compounds for further development. This approach has not been very successful in the development of new classes of antibiotics, and an issue that has been recognised is the lack of chemical diversity in compound libraries (Payne *et al.,* 2007; Tommasi *et al.,* 2015). Attempts have been made to improve libraries with broader chemical space representatives (Gwynn *et al.,* 2010; Silver, 2011; Chopra, 2013). These improvements include the incorporation of new scaffolds with different chemical properties, and the development of pathway-specific libraries. Another approach is the search for a lead compound using fragment screens (Erlanson, 2006). In a more sophisticated way, phage display (Løset and Sandlie, 2012), and DNA-encoded libraries (Franzini *et al.,* 2014) can also be used. The former has been used for creating biomacromolecule libraries (*e.g.* peptides, proteins, antibodies and nucleic acids). DNA-encoded libraries present a new way of synthesis and display of organic compound libraries. Lead compounds have been identified using DNA-encoded libraries for various eukaryotic targets (Franzini *et al.,* 2014), however in this approach a pool of different compounds are tested at once and this hinders the identification of the hits. Compound, fragment, phage display and DNA-encoded libraries can be screened against the whole-cell (phenotypic screen), an approach that might inform about which compounds can enter the cell. However, the identification of the drug target(s) is challenging and without this information, the optimisation of the compounds could be difficult and more time-consuming than in the single-target approach.

Despite extensive HTP screening against specific targets, only a few leads have been identified for further development. This has lead to questions on how appropriate the compound libraries are for antibacterial drug discovery (Payne *et al.,* 2007; Gwynn *et al.,* 2010). The failure rate in the lead compound search has been assessed by GSK

(GlaxoSmithKline, Payne *et al.,* 2007) and AstraZeneca (Tommasi *et al.,* 2015) using different libraries and targets. The GSK program obtained 5 lead compounds from 70 screening campaigns carried out between 1995 and 2001. Payne *et al.,* (2007) suggested that their library was not appropriate for antibacterial development and the use of more chemically diverse compounds might be an improvement. AstraZeneca reported results of 65 screening campaigns (between 2001 and 2010), containing 8 targets in common with the GSK program. Fifteen lead compounds were obtained, however further investigations have not resulted in potential drugs, showing problems in terms of cell penetration (Tommasi *et al.,* 2015).

The first leads provide chemical information on what might form the basis for drug discovery. To progress to the synthesis of active compounds it is important to have accurate structural data. Target structures, including apo-form and with relevant ligands (*e.g.* substrates) or lead compounds, give information about binding modes as well as the physical and chemical properties of the ligand environment providing data for lead optimisation and rational drug design. In addition, protein-ligand complexes contribute to validation of hits obtained from HTP assays. In theory, it is possible to use structures as templates for developing *in silico* inhibitors. Such information can be used to design specific single-target based libraries (Frearson *et al.,* 2007; Gwynn *et al.,* 2010). However, these compounds usually do not support translation to the whole-cell environment as they are not able to overcome the membrane barriers of the microorganism (Gwynn *et al.,* 2010).

It has been recognised that directing drug discovery towards a single target might not be the most appropriate strategy for developing antibiotics (Gwyn *et al.,* 2010). A more promiscuous drug would, in principle, require a more complex resistance system.

Indeed, the already clinically implemented single-target drugs (*e.g.* fosfomycin, fusidic acid and rifampicin) have shown a higher tendency to generate resistance than antibiotics targeting several proteins (*e.g.* β-lactams, fluoroquinolones and vancomycin) (Gwynn *et al.,* 2010; Silver, 2012). Combinatorial therapies, which combine various antibiotics to circumvent issues of drug resistance, are already in use. This is the case for treatment of infections caused by *Plasmodium falciparum* (the causative agent of malaria) and *Mycobacterium tuberculosis* (Fischbach, 2011; Tamma *et al.,* 2012). Multidrug resistance in Gram-negative bacteria has been attributed to the drug extrusion activity of efflux pumps in many clinical isolates (Li *et al.,* 2015; Chang *et al.,* 2015). In this context, treatment using combinatorial therapies could include cytosolic-target drugs, newly developed or already approved, plus an inhibitor of the efflux pump systems. This could increase susceptibility and toxicity to the cytosolic drug and could be used to reverse antibiotic resistance (Lomovskaya and Watkins, 2001).

Alternative strategies to conventional drug discovery have been proposed. These include phage therapy (Nobrega *et al.,* 2015) and RNA interference-based therapy (Dyawanapelly *et al.,* 2014). Despite the promise of such alternative treatments, there has been little progress. Conventional approaches are still considered the most likely to progress development to antibacterial drugs (Gwynn *et al.,* 2010).

1.2.2.3 Target selection approaches

Genetic and chemical data are required to assess the potential of antibacterial targets (Monaghan and Barret, 2006; Frearson *et al.,* 2007). To handle the large amount of information present in various databases, different target selection strategies can be combined to generate a list of candidates (Frearson *et al.,* 2007). A common approach is the prediction or analysis of gene essentiality (Freiberg and Brötz-Oesterhelt, 2005).

This can either be carried out experimentally or inferred from phylogenetically related bacteria (White and Kell, 2004; Doyle *et al.,* 2010; Monye *et al.,* 2013). When genome-scale transposon knock-outs are available this information is used to generate a probable essential genes list (Moynie *et al.,* 2013). Another approach is to exploit known antibacterial drug targets in the bacteria of interest (Frearson *et al.,* 2007; Doyle *et al.,* 2010), so that these proteins and the pathways to which they belong can became targets. Such approaches generate a large number of potential candidates that need to be further analysed and filtered. Various criteria are generally used to conduct the selection and prioritisation of potential antibacterial candidates including precedence, essentiality, druggability, structural biology feasibility, selectivity, assay feasibility and spectrum of action (Frearson *et al.,* 2007).

Druggability is an important feature for target selection, corresponding to the propensity of drug-like compounds to interact with the target. It can be analysed from the structure or from large-scale bioactivity databases. Drug-likeness of a compound was initially assessed by Lipinski's Rule of Five (Ro5, Keller *et al.,* 2006), which was derived from empirical data on oral neuroactive compounds. The Ro5 suggests that, for the purpose of bioavailability and usefulness in subsequent chemical modifications, drug-like compounds should have: 1) no more than 5 hydrogen bond donors, 2) no more than 10 hydrogen bond acceptors, 3) molecular weight less than 500 Da and 4) an octanol-water partition coefficient (logP) below 5. More recently, the Quantitative Estimate of Drug-likeness (QED) was established as an improvement of the Ro5 and to allow the assessment of large-scale chemogenomic data (Bickerton *et al.,* 2012). This is based on eight criteria: molecular weight, octanol-water partition coefficient, molecular polar surface area, number of hydrogen bond donors, hydrogen bond acceptors, rotatable bonds, aromatic rings and structural alerts. The QED value ranges from 0 (non-drug-

like) to 1 (drug-like). A protein target is considered druggable when it is able to bind drug-like compounds. In this case, QED values of published active compounds in ChEMBL (www.ebi.ac.uk, Bento *et al.,* 2014) and the number of drug-like compounds for each target can be used for estimating target druggability. Additionally, if a protein structure is available, druggability can be further evaluated using software such as DrugPred (Krasowski *et al.,* 2011) or the Molsoft ICM pocket finder tool (www.molsoft.com). These programs generate a drug-like density score (DLID, Sheridan *et al.,* 2010) that considers volume, buriedness and hydrophobicity of the pockets. The DLID values range from -3 to 2 and a pocket is considered druggable when a value equal to or greater than 0.5 is obtained.

A structure can inform about active site pocket properties, substrates or inhibitor binding, and therefore, the reaction mechanism. This can be used for structure based drug design and for validation of drug binding. Feasibility of the targets for crystallisation can be predicted using XtalPred (Slabinski *et al*., 2007). A crystallisation probability score is calculated using various criteria (protein length and molecular weight, pI, instability index, content of Cys/Met/Trp/Tyr, percentage of coiled-coil or membrane regions among others) and compared with data at TargetDB (which also contains failure crystallisation data, www.targetdb.pdb.org, Chen *et al.,* 2004). For each target, structures of homologues in the PDB (Protein Data Bank, www.rcsb.org) are identified and considered. The presence of ligand-bound and high-resolution structures will inform about the feasibility of obtaining good-quality crystals to conduct inhibitor-bound studies.

Another feature to evaluate when considering antibacterials is the selectivity of the target between the pathogen and the host. Ideally, the target should provide

discrimination to selectively target the bacteria. This can be achieved by lack of an homologue in the host, binding site selectivity (different residues in the binding site pockets), differences in the kinetics and binding mechanisms, the target essentiality in bacteria and host, and differences in drug transport and efflux systems (Huggins *et al.,* 2012; Torok *et al.,* 2012). Also, depending on the strategy chosen (broad-spectrum or selective targets), the presence of orthologues in other bacteria can be included in the criteria (spectrum of activity) (Frearson *et al.,* 2007). Finally, information in the literature and the BRENDA database (Chang *et al.,* 2015) concerning assay feasibility, the existence of HTP assays, and inhibitors are also considered. Two approaches are generally followed in the search for inhibitors (Gwynn *et al.,* 2010; Silver, 2011; Chopra, 2013). On one hand, rational design for new inhibitors based on the reaction mechanism and the available structures are often used. To validate the compounds an activity assay is required. On the other hand, to investigate novel inhibitor scaffolds, compound library screens can be carried out and this may require the development of appropriate assays.

1.2.3 The future

Problems in the current approaches and libraries have been identified and issues are being addressed (Silver, 2012; Chopra, 2013; Tommasi *et al.,* 2015). Improvements in compound library collections are being implemented (Silver, 2012; Tommasi *et al.,* 2015) and may result in more efficient lead searches. In this respect, the identification of new chemical classes (Davis *et al.,* 2015; Hamamoto *et al.,* 2015; Iscla *et al.,* 2015; Ling *et al*., 2015) provides optimism to exit the Void Age. However, campaigns for social awareness to promote the correct use of antibiotics and, therefore, avoid their rapid failure should be carried out along with the development of new antibacterial

drugs (Chopra, 2013). Additionally, research to develop alternative therapies may help to diversify strategies against resistant bacteria infections (Chopra, 2013).

1.3 Biowarfare agents

Some bacteria and viruses have been used as biowarfare agents, or bioweapons (Barras and Greub, 2014). Among these are the pathogens causing anthrax, glanders, plague and tularaemia. The critical situation in antibacterial resistance intensifies the concern about the employment of multidrug resistance agents for criminal acts. The U.S. Centers for Disease Control and Prevention (CDC, www.cdc.gov) have categorised biowarfare agents into different priority groups. Category A agents can be easily spread within the population, cause disease with high mortality rates and generate panic. Examples would be *Bacillus anthracis*, *Francisella tularensis* and *Yersinia pestis*. Agents from Category B spread less easily than Category A agents and have lower mortality rates. A representative of these agents is *Burkholderia pseudomallei*. Category C agents are characterised as pathogens that could be engineered for mass spread, are easy to obtain and produce, and present high morbidity and mortality rates. Multidrug-resistant *Mycobacterium tuberculosis* is an example of a Category C agent (Anderson *et al.,* 2012). Characteristics of the Category A and B organisms mentioned above are summarised in Table 1.1.

	B. anthracis (Anthrax)[1,2]	*B. pseudomallei* (Melioidosis)[1,2,5]	*F. tularensis* (Tularemia)[1,2,6]	*Y. pestis* (Plague)[1,2,3,4]
Organism	Gram-positive, extracellular, endospore forming, presents a capsid. It can survive in aerobiosis and anaerobiosis.	Gram-negative, intracellular, aerobic, motile outside the host cell.	Gram-negative, cocobacillus, aerobic, facultative intracellular, non-motile, infects macrophages.	Gram-negative, intracellular, facultative anaerobic, motile outside the host cell.
Genome organisation	One chromosome and two virulence plasmids.	Two chromosomes containing nine different secretion systems that confer intrinsic multidrug resistance.	One chromosome and one plasmid.	One chromosome and several virulence plasmids.
Life cycle	Zoonic. Transmission by contaminated soil, herbivores, wool, and human-to-human.	Transmission by contaminated soil and water.	Zoonic. Transmission by contaminated soil, water, food, ticks, flies, mosquitoes, rodents and human-to-human.	Zoonic. Transmission by flea bites, rodents, and human-to-human.
Drugs and vaccines	Ciprofloxacin and doxycycline. There is a vaccine for military purposes.	Amoxicillin, tetracycline, trimethoprim-sulfamethoxazole.	Gentamicin, streptomycin, tetracycline and chloramphenicol.	Streptomycin, gentamicin, doxycycline, ciprofloxacin and chloramphenicol.
Resistance observed	Multidrug resistance.	Intrinsic multidrug resistance.	Streptomycin resistance.	Multidrug resistance.
Geographic distribution	20,000-100,000 cases per year: Middle East, Indian subcontinental, Asia, Africa and Latin America.	Endemic in tropical and subtropical regions of southeast Asia and Northern Australia.	Worldwide.	2,500 in the world. Africa (about 76% of the cases), Asia. USA: 10 cases per year (Southwest). It is considered a re-emerging disease.

Table 1.1: Summary of *B. anthracis*, *Y. pestis*, *B. pseudomallei* and *F. tularensis* characteristics. ([1]) Greenfield *et al.,* 2002. ([2]) Pohanka and Skladal, 2009. ([3]) Galimand *et al.,* 2006. ([4]) Anisimov and Amoako, 2006. ([5]) White, 2003. ([6]) Carvalho *et al.,* 2014.

1.4 Aim and objectives

Together with the Dstl (Defence Science and Technology Laboratory) this project is designed to address the evaluation and initial validation of new antibacterial drug targets for the biowarfare agents *B. anthracis*, *B. psesudomallei*, *F. tularensis* and *Y. pestis*. To achieve this, the following objectives were defined:

1. Selection of potential targets and prioritisation: Generate a list of candidates taking into account information in databases, the literature and additional data from the Dstl.
2. Genetic validation of the targets: Confirmation of gene essentiality for cell growth or viability. This is being carried out by the Dstl.
3. Early-stage chemical validation. This requires:
   - Recombinant protein production to support the program.
   - Structural characterisation of the potential targets using crystallography.
   - Biochemical characterisation/validation of the targets using different techniques for determining enzymatic parameters, ligand binding constants and their performance in chemical library screening campaigns.

PART TWO

SELECTION AND PRIORITISATION OF POTENTIAL TARGETS

In consultation with the Dstl (Defence Science and Technology Laboratory), ten potential targets in *Bacillus anthracis*, *Burkholderia pseudomallei*, *Francisella tularensis* and *Yersinia pestis* were selected based on the criteria detailed in section 1.2.2.2 (precedence, essentiality, druggability, structural biology feasibility, selectivity and assay feasibility). In this chapter, the decisions that guided target selection and prioritisation are explained.

2.1 Target selection strategy

2.1.1 Candidates for target selection

The first approach to generate a list of target candidates was based on transposon-directed insertion site sequencing (TraDIS, van Opijnen and Camilli, 2013). Genome-scale transposon knock-outs (TKO) of *Y. pestis* was carried out by the Dstl team. TKO clones showing defects in cell growth were selected and sequenced using TraDIS. The results provided information about potential essentiality of groups of genes, consequently, false positives are expected from this approach and for selected targets

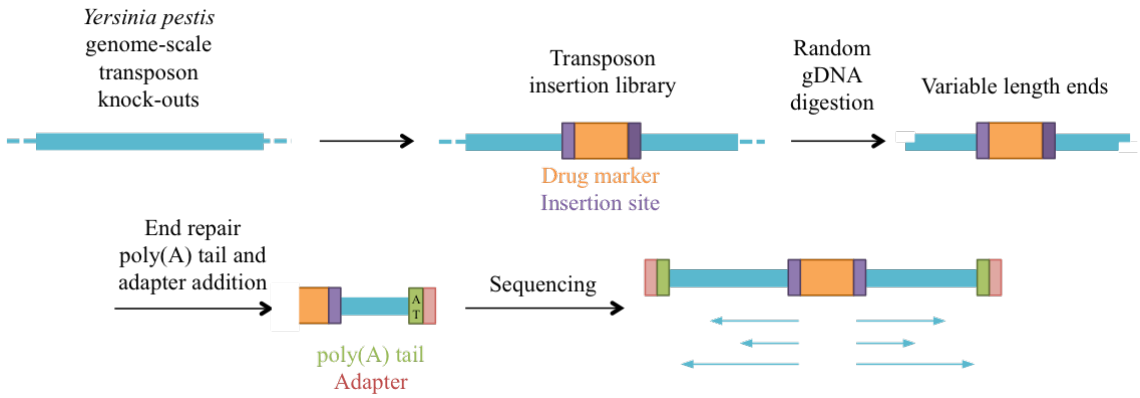


Figure 2.1: Schematic of the TraDIS method adapted from Opijnen and Camilli (2013).

direct genetic validation is then required. To assist the completion of TraDIS, it is necessary to produce a transposon insertion library. The genomic DNA (gDNA) is digested with random DNA restriction enzymes generating products carrying the transposon, used as the sequencing starting point, and variable-length ends containing gDNA. The DNA ends are repaired, a poly(A) tail added and an adaptor sequence included to facilitate sequencing (Figure 2.1). Sequenced data were compared to the wild-type *Y. pestis* gDNA and knocked-out genes identified to generate a list of potential essential genes (Appendix A).

2.1.2 Assessment of selection criteria

For the selection of targets for *B. anthracis*, *B. pseudomallei* and *Y. pestis* different datatabases were used. Firstly, AEROPATH (www.aeropath.eu) gave information about gene essentiality, druggability, presence of homologues in the host and/or other Gram-negative bacteria, crystallisation feasibility, and published drug-like compounds that interact with the targets. AEROPATH is a database prepared for target assessment in *Pseudomonas aeruginosa* and was developed at the University of Dundee. Essentiality information in AEROPATH and DEG (Database of Essential Genes, www.tubic.tju.edu.cn/deg, Zhang *et al.,* 2004) databases was compared to the results from the TraDIS data for essentiality cross-validation. On one hand, DEG showed there are genes characterised as essential for at least one bacterial organism. On the other hand, the AEROPATH database reported some of the potential targets were not essential for *P. aeruginosa*. Information about drug-like compounds was further studied in the ChEMBL database (Bento *et al.,* 2014), which contains a summary of the medicinal chemistry literature, and seeks to link chemical structures with specific targets. Previous work at Dstl and Prof. William N. Hunter labs was also used to inform target selection. The consistency in prediction of essentiality among different

organisms, presence of an available enzymatic assay, accurate structure data, and the existence of human homologues with high sequence identity were considered followed by the analysis of characterised inhibitors and QED values. Some enzymes involved in DNA/RNA biosynthesis, ribosome constituents, the ATP synthase system, scaffold or non-catalytic and uncharacterised proteins were not considered. Additionally, proteins with a high sequence homology (more than 50%) with human homologues were discarded. Four potential targets had a HTP assay described (marked as green in the traffic light score, see Appendix A). These are acetyl-CoA carboxylase:carboxyltransferase subunit α (AccA), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE) and UDP-N-acetylglucosamine acyltransferase (LpxA) (Appendix A). However, there is no evidence of lead compound follow up so this could indicate difficulties and they were not included for the final target selection.

2.1.2.1 Traffic light score

The collected information was classified using the traffic light score defined in Frearson *et al.,* (2007). For this project, variations in the definitions were implemented and are shown in Table 2.1. For evaluation of essentiality, data from TraDIS together with the information in AEROPATH and DEG databases were considered. Druggability was evaluated with the QED score and the presence of drug-like data separately. The first one was marked with a colour scale; red lowest QED score and green as the highest score, and the second as indicated in Table 2.1. For “human homologs”, due to the large number of targets, a first check was pursued using the sequence identity level and a further analysis was carried out for the potential targets looking for evidence of selectivity or essentiality in humans (see section 2.2 for details). The analysis of the tractability as a target of the candidates was done checking the size and subcellular

location (not coloured), the availability of assays and structures, and XtalPred prediction. The last was marked with a colour scale; red showing the highest scores and green the lowest.

Criterion		Red	Amber	Green
Essentiality	AEROPATH[1]	Non-essential. No evidence for gene essentiality	Probably essential. There is weak evidence of gene essentiality	Essential. Strong genetic evidence of gene essentiality for cell growth
	DEG[2]			
Close human homologues[1]		Human homologues present, showing 30% sequence identity	Human homologues present, sequence identity lower than 30%	No human homologues
Druggability	Drug-like CPDs[1,4,5]	No drug-like inhibitors	Known drug-like inhibitors in other bacteria	Known drug-like inhibitors in any of the targeted-bacteria[1,3,4] and ChEMBL homologues[1,2]
	ChEMBL homologues[1,3]	No ChEMBL homologues	ChEMBL homologues	The target has active compounds in ChEMBL (so-called ChEMBL target)
Tractability as a target	Assay feasibility[4,5]	An assay has not been developed or involves radioactivity.	An assay has been described (non-HTP assay)	HTP assay available
	Structural information[1,6]	No structures (including other bacteria)	Structures available but not for the targeted organisms	Structures available for the targeted organisms

Table 2.1: Traffic light score definitions for criteria classification. CPDs: compounds abbreviation. Information collected from ([1]) AEROPATH, ([2]) DEG, ([3]) ChEMBL, ([4]), BRENDA database, ([5]) Literature, and ([6]) PDB database.

2.2 Selected targets

From 3886 genes analysed in the TraDIS data, 130 were selected as encoding for probable essential targets. This were filtered as indicated in section 2.2 and resulted in 75 candidates for further analysis (Appendix A). Prioritisation of the targets was

carried out accordingly to the experimental results and the progress towards target validation.

2.2.1 First selected targets

The first targets selected were kynurenine formamidase (KynB, Zummo *et al.,* 2012), dihydrofolate synthase:folyl-poly-glutamate synthase (bifunctional protein FolC, Sheng *et al.,* 2002), 3-hydroxyacyl-[acyl-carrier-protein] dehydratase (FabZ, White *et al*, 2005), a lipid A export ATP-binding permease (MsbA, Ghanei *et al.,* 2007) and AdoMet-dependent 5-methyluridine methyltransferase (RsmH or MraW, Wei *et al.,* 2012). KynB is the second enzyme involved in the tryptophan degradation pathway, the bifunctional protein FolC belongs to the folyl-poly-glutamate biosynthesis pathway, FabZ and MsbA are required for maintaining the integrity of the cell membrane, and RsmH is implicated in the post-translational modifications of ribosome RNA. FolC and MsbA have human homologues. In the first case it is referred to as the mitochondrial FolC, which presents only one of the activities (folyl-poly-glutamate synthase), and selectivity towards the bacterial FolC might be feasible (Wang *et al.,* 2010). MsbA homologs involve proteins from the ABC transporter family and, as expected, the ATPase domain is conserved in this family. However, selectivity might be possible targeting other sites on the protein. Structures, inhibitors and enzymatic assays have been described for FabZ, FolC and MsbA. Less information is available for KynB and RsmH for which no inhibitors have been described. Additionally, no structural information was available for KynB, however assays have been described in both cases and XtalPred scores suggested crystallisation is feasible. Selection of KynB and RsmH was based on previous results; Dr. V. Srikannathasan, in Prof. Hunter lab, showed KynB has a different fold to the eukaryote homologue and it was predicted as essential in *P. aeruginosa* from genome-scale transposon experiments (Jacobs *et al.,* 2003;

Liberati *et al.,* 2006), and the Dstl labs had essentiality data for RsmH. These two targets were not identified from the TraDIS results. Experimental work towards the chemical validation of these first candidates was initiated and resulted in the prioritisation of KynB. In parallel, genetic validation, involving conditional knock-out generation for the different bacteria as well as random mutations, was being carried out at Dstl. Before moving to high-throughput screening to search for KynB inhibitors, Dstl reported KynB as non-essential for *B. anthracis* or *B. pseudomallei* growth, but indicated it might be involved in *B. pseudomallei* virulence (Prof. Richard W. Titball, personal communication). New information available (see Part Three and Appendix C) suggested FabZ, KynB and RsmH might not be suitable targets, so a further selection was carried out.

2.2.2 Second selection

The second selection included D-alanine—D-alanine ligase (Ddl, Prosser and de Carvalho, 2013), the cell division protein FtsZ (Errington *et al.,* 2003), CTP synthase (PyrG, Yoshida *et al.,* 2012), the proteolytic component of the caseinolytic protease (ClpP, Brötz-Oesterhelt and Sass, 2014), and the efflux pump NorM (Su *et al.,* 2008), for further assessment. Ddl is involved in peptidoglycan layer biosynthesis, a reference pathway (Bermingham and Derrick, 2002; Bugg *et al.,* 2011), together with folate metabolism, in drug discovery. Structures were available (Table 2.2), inhibitors identified and an enzymatic assay described suggesting this protein was a good candidate. In the case of FtsZ, structure and inhibitors were described however an enzymatic assay needs to be developed. PyrG is present in humans but is not essential and is not the preferred route for CTP synthesis in humans (Huang and Graves, 2002), so selectivity might be viable. There are examples of inhibitors, structures and assays available for this target. Also, in this second selection, there were two proteins that

could be used alternative approaches. ClpP produces toxicity to the cell when it is overexpressed or overactivated due to unspecific proteolysis and thus cell death. In this context, ClpP was considered. There is a human homologue (mitochondrial ClpP) but it might be possible to get selectivity towards bacterial proteins (Compton *et al.,* 2013). Drug efflux pumps, such as NorM, produce intrinsic drug resistance. *B. pseudomallei* has a big drug extrusion network that confers multidrug resistance (Schweizer, 2003; Biot *et al.,* 2011). Therefore, efflux pumps are essential for the growth of this bacterium under antibacterial drug stress, and inhibitors of efflux transport could be used to intensify the effect, or reverse resistance, of already known antibiotics in combinatorial therapies (Lomovskaya and Watkins, 2001).

Yp gene name	Essentiality		Human homologue	Druggability			Tractability as a target				
	Pa	DEG		QED	ChEMBL	Drug-like compounds	Size (AA)	Location	Assay	Structures (resolution)	Xtal-Pred
kynB	P	E	No	-	0 / 0	None	213	Cytoplasm	SA (Katz *et al.*, 1987)	No KynB structure but there is a structural homologue (1R61)	2
folC	E	E	Yes, 23%	0.27	2 / 0	Two inhibitors for malaria (Wang *et al.*, 2010)	429	Cytoplasm	SA (Bognar *et al.*, 1985)	15 structures, *Yp* (1.5 - 2.5 Å)	2
fabZ	E	E	No	0.53	1 / 0	Inhibitors of *Mt* (μM) (He *et al.*, 2009)	146	Cytoplasm	SA (He *et al.*, 2009)	19 structures (2.0 - 3.0 Å)	2
rsmH*	-	E	No	-	0 / 0	None	313	Cytoplasm	SA (Carrión *et al.*, 1999)	4 structures (1.5 - 2.5 Å)	3
msbA	E	E	Yes, 38% [A]	0.8	52 / 0	Inhibitors for *Ec* (no K_i values) [B]	603	Inner membrane	Calorimetry (Ghanei *et al.*, 2007)	8 structures (2.5 - 5.5 Å)	5
ftsZ	E	E	No	0.57	5 / 1	Inhibitors for *Bt* [B]	394	Cytoplasm	Radioactivity (Lu *et al.*, 1998), GTPase activity (Lu *et al.*, 2001)	63 structures (1.2 - 4.0 Å)	5
ddl	N	E	No	-	2 / 1	Inhibitors *Ec*, *Hp* (nM) (Wu *et al.*, 2008)	319	Cytoplasm	SA-coupled (Wu *et al.*, 2008)	29 structures, *Ba*, *Bp*, *Yp* (1.5 - 3.0 Å)	2
pyrG	E	E	Yes, 45% [A]	-	0 / 0	Inhibitors for *Ec* (mM/nM) [B]	542	Cytoplasm	SA (Lunn and Bearne, 2004)	8 structures (1.5 - 3.5 Å)	5
norM	N	N	No	-	0 / 1	Inhibitors (μM) (Long *et al.*, 2008)	457	Inner membrane	Various transport assays (Long *et al.*, 2008)	6 structures (3.4 - 4.2 Å)	5
clpP	N	E	Yes, 50% [A]	-	0 / 2	Activators (Brotz-Oesterhelt, 2014)	213	Cytoplasm	Fluorescence-based protein degradation assays (Brotz-Oesterhelt and Sass, 2014)	39 structures, *Ft* (1.5 - 3.5 Å)	1

Table 2.2: Target selection criteria and traffic light scoring. Abbreviations are: *B. anthracis* (*Ba*), *B. pseudomallei* (*Bp*), *P. aeruginosa* (*Pa*), *Y. pestis* (*Yp*), *E. coli* (*Ec*), *Mycobacterium tuberculosis* (*Mt*), *Bos taurus* (*Bt*) and *Helicobacter pylori* (*Hp*). (*) Not characterised in *P. aeruginosa*, so information corresponds to the *E. coli* homologue. ([A]) Coloured in amber because there is some indication of selectivity between the bacteria and the host. ([B]) Obtained from BRENDA database. E is essential, P probable essential and N non-essential. In the ChEMBL column the first values corresponds to the number of ChEMBL homologues and the second the number of ChEMBL targets. The size column indicates the number of amino acids (AA) of the *P. aeruginosa* protein.

2.3 Target prioritisation

Selected target criteria analysis and results of the Dstl genetic validation are collected in Table 2.2 and the target prioritisation decisions shown in Figure 2.2. Preliminary data from Dr V. Srikannathasan served to prioritise KynB studies. Results from the Dstl on KynB and FolC essentiality, and problems with expression and purification of FabZ and RsmH, set FolC as the new priority. Later, failure on the optimisation of the previously described FolC enzymatic assay and the successful structure characterisation of Ddl, promoted the second to highest priority. Reprioritisation of FolC resulted from the failure of screening campaigns against Ddl and the identification of a new FolC assay. The following chapters (Part Three, Four, Five, Six and Appendix C) will include the experimental data obtained during the assessment of these potential antibacterial drug targets.

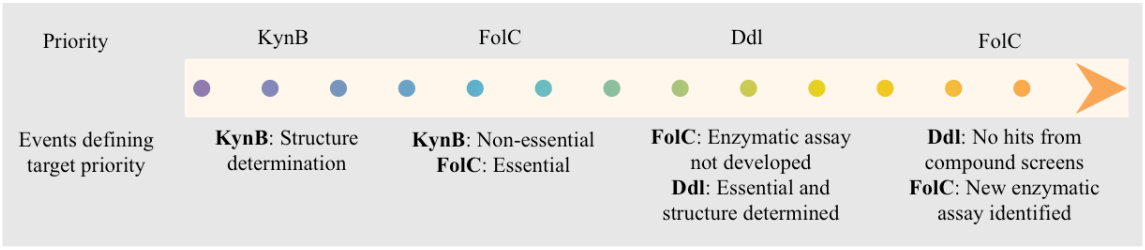


Figure 2.2: Prioritisation of the targets during the course of the project. A summary of the events defining target priority is included. KynB was first prioritised regarding preliminary structural data. Essentiality data from the Dstl labs (and collaborators) involved the prioritisation of FolC and work on KynB was not followed. Then, problems on the FolC assay optimisation and Ddl essentiality data from the Dstl, suggested the prioritisation of Ddl. Finally, due to the lack of Ddl inhibitory hits from the compound screening campaigns and the identification of a new assay for FolC, the latter was incorporated as a priority.

PART THREE

KYNURENINE FORMAMIDASE

3.1 Introduction

Tryptophan degradation is important both in prokaryotes and eukaryotes for the production of anthranilate and the precursor for NAD biosynthesis, quinolinate (Kurnasov *et al.,* 2003). Anthranilate is a metabolite that supports the synthesis of aromatic amino acids (Voet and Voet, 2004), and is the main constituent of chorismate, which is used in the formation of the folate precursor *para*-aminobenzoate (Satoh *et al*., 2014). In bacteria, tryptophan is also used in antibiotic production and provides a source of carbon and nitrogen (Kurnasov *et al.,* 2003).

Eukaryotic tryptophan catabolism products L-kynurenine and quinolinate are involved in neurogenic disorders (Stone and Darlington, 2002). The enzyme kynurenine transaminase (Enzyme Commission number (EC) 2.6.1.7, Stone and Darlington, 2002) produces kynurenic acid, an antagonist of acetylcholine receptors (AChRs), from L-kynurenine. Kynurenic acid can inhibit *N*-methyl-D-aspartate (NMDA) glutamate receptors (NMDARs) but it has been shown it presents a greater affinity for AChRs. On the other hand, quinolinate is an agonist of NMDARs. Either over-activation or inhibition of these receptors causes neuronal cell damage. In Alzheimer's disease, a high concentration of kynurenic acid in the brain hinders memory and learning processes due to the inhibition of AChRs. Quinolinate activates NMDARs producing cell damage by reactive oxygen species formation. It also promotes the secretion of cytokines and chemokines by astrocytes enhancing or activating inflammatory responses. Quinolinate has been related to the inhibition of phosphate removal from tau

triggering the increase in poly-phosphorylated tau concentration forming amyloid plaques (Rahman *et al.,* 2009), a characteristic feature of Alzheimer's disease. In Huntington's, this pathway is also altered and shows a low concentration of L-kynurenine and high levels of quinolinate in the brain which activates the overexpression of huntingtin contributing to the formation of toxic aggregates (Maddison and Giogini, 2015). Kynurenic acid seems to be involved in schizophrenia, and other psychiatric disorders, by inhibiting AChRs. Agonists of these receptors can be used to reduce the symptoms of these illnesses (Stone *et al.,* 2013; Stone and Darlington, 2002).

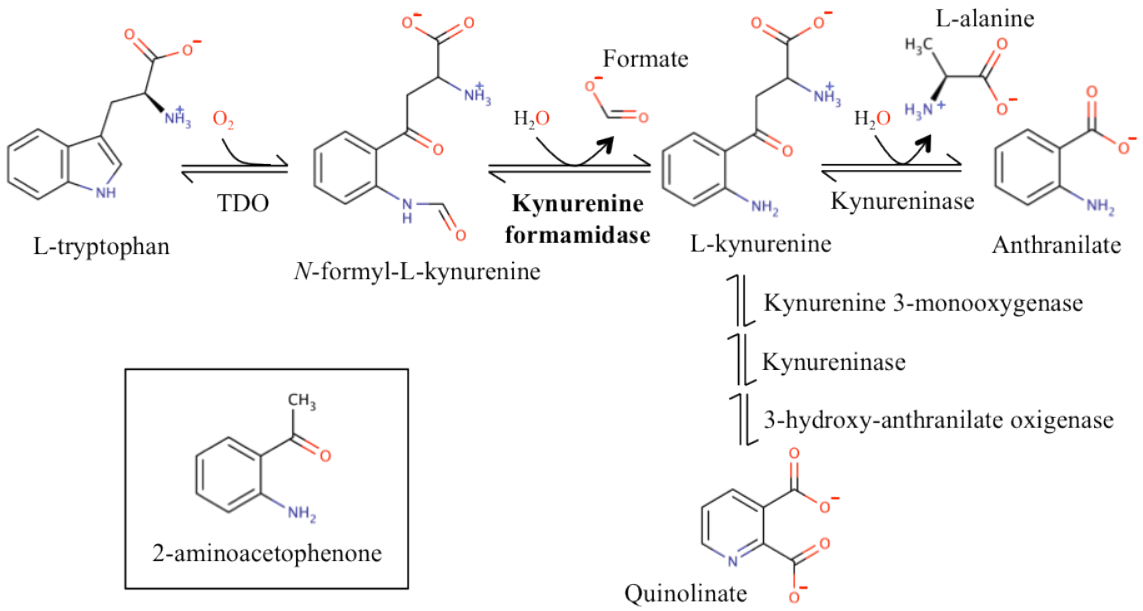


Figure 3.1: Tryptophan degradation pathway via L-kynurenine. In bold, the protein of interest. Three enzymes are needed for the conversion of L-kynurenine to quinolinate. In the box, 2-aminoacetophenone, which shares structural features with L-kynurenine.

The aerobic degradation pathway via anthranilate is the preferred route in tryptophan catabolism and involves three enzymes; tryptophan 2,3-dioxygenase (TDO, EC 1.13.11.11), kynurenine formamidase (EC 3.5.1.9) and kynureninase (EC 3.7.1.3) (Figure 3.1). The first enzyme performs the oxidation of L-tryptophan to *N*-formyl-L-kynurenine (NFK), then kynurenine formamidase hydrolyses the amide group of NFK

to form L-kynurenine and formate. The third enzyme, kynureninase, is then responsible for anthranilate formation (Kurnasov *et al.,* 2003). L-kynurenine can also progress into the quinolinate pathway to support NAD biosynthesis. The constituent enzymes of the degradation pathway are highly conserved with the exception of kynurenine formamidase (Zummo *et al.,* 2012). The eukaryotic formamidase, called KFase, has been intensively studied because it is implicated in neurogenic disorders, however, less work has been carried out in the prokaryote homologue, called KynB (Han *et al.,* 2012; Stone *et al.,* 2013). This enzyme has been annotated in the AEROPATH database (www.aeropath.lifesci.dundee.ac.uk, Moynie *et al.,* 2013) as essential for *Pseudomonas aeruginosa* growth and consequently it was included in the program for further characterisation. Comparing the amino acid sequences, KFase and KynB appear to be unrelated proteins as will be explained later.

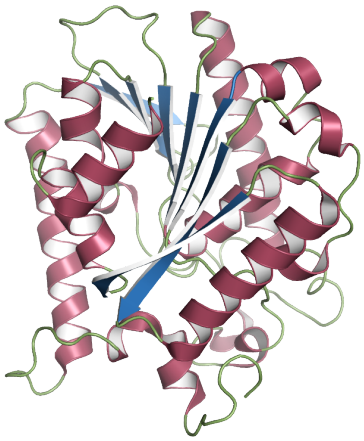


Figure 3.2: Cartoon representation of *Dm*KFase (PDB code 4E11).

The amidase reaction can occur with either acidic or basic hydrolysis (Fyfe *et al.,* 2008). In the latter case, a reactive cysteine or serine is responsible for catalysis. The known KFase active site presents a catalytic triad comprising a serine, an aspartate and a histidine. The reactive serine is located in a highly conserved amino acid sequence characteristic of the serine hydrolase family (Gly-X-**Ser-**X-Gly). Other known amidases such as trypanothione synthetase amidase (Fyfe *et al.,* 2008), nicotinamidase (Fyfe *et al.,* 2009) and the nitrilases (Pace and Brenner, 2001) use a reactive cysteine to perform the reaction. The structures of *Drosophila melanogaster* (*Dm*KFase, Figure 3.2) and *Saccharomyces cerevisiae* KFase (*Sc*KFase) revealed a similar α/β hydrolase fold (Han *et al.,* 2012; Wogulis *et al.,* 2008) even though their sequences share only 12% identity. This fold is characterised by eight

parallel β-strands surrounded by 13 α-helices. There exist structural similarities between KFase, carboxylesterases (Wogulis *et al.,* 2008) and 5'-pyridoxyl phosphate-dependent kynurenine aminotransferase (Rossi *et al.,* 2004).

KynB and KFase have a low amino acid sequence identity (10%) with a lack of a conserved motif. Comparing the KynB secondary structure prediction with the KFase structure, KynB seemed to present a distinct fold. Additionally, the size is considerably different with 300 amino acids in the case of KFase and 213 for KynB. All these factors suggest KynB is structurally and mechanistically unrelated to KFase. These observations also apply to other known amidases or amidohydrolases indicating KynB might also be different from these enzymes. Based on this, and that KynB has been described as an essential protein (AEROPATH; Farrow and Pesci, 2007), this enzyme was considered a potential antibacterial drug target. KynB from *B. anthracis* (*Ba*KynB), *B. cenocepacia* (*Bc*KynB) and *P. aeruginosa* (*Pa*KynB) were investigated. The work was focused on the structure and biochemical characterisation of these proteins. To get information about the reaction mechanism and binding of the substrates, co-crystallisation with L-kynurenine and the compound 2-aminoacetophenone was investigated. Additionally, confirmation of the kynurenine formamidase activity in these proteins was required. Once structures were obtained and the enzymatic assay optimised, the initial chemical validation of *Ba*KynB as a new antibacterial drug would follow. However, results from the Dstl showed this protein is not essential for *B. anthracis* hence no further validation work was pursued. In this chapter, structures of apo-KynB and *Ba*KynB complexed with 2-aminoacetophonone are reported. The structures reveal a distinct fold from other amidases and give information about important features for molecular recognition. Moreover, the *Ba*KynB structure

complexed with 2-aminoacetophoneone gave information about the reaction mechanism.

3.2 Experimental procedures

Most of the genes reported in this thesis have been cloned and the encoded proteins purified using similar protocols. General protocols are described here and variations will be noted as appropriate.

3.2.1 Recombinant protein production

Synthetic genes encoding KynB in *P. aeruginosa* (*Pa*KynB, UniProt: Q9I234), *B. cenocepacia* (*Bc*KynB, B4E9I9) and *B. anthracis* (*Ba*KynB, Q81PP9) were purchased (Genscript) having been codon optimised for expression in *Escherichia coli* K12. They were cloned into a modified pET15b vector (Novagen, called pET15bTEV) to create plasmids that produce an N-terminal hexahistidine-tagged (His_6-tag) protein and a tobacco etch virus (TEV) protease cleavage site. The restriction enzymes used were NdeI (at the 3' end of the DNA strand) and BamHI (at the 5' end). All constructs were sequenced to check their integrity by the sequencing facility at the University of Dundee. Cloning of *Pa*KynB was performed by Dr. V. Srikannathasan and *Bc*KynB by Dr. M. Zoltner (see Table 3.1 for variations in the protocol). No mutations were observed in any of the genes.

Gene expression and purification of the proteins started with freshly transformed *E. coli* strains, generally BL21(DE3). Bacteria were cultured at 37ºC in 10 ml of LB broth containing 50 μg ml^{-1} carbenicillin. In the case of pLysS strains, 20 μg ml^{-1} chloramphenicol was also present in combination with carbenicillin. These cultures

were used as inoculum for 1 litre of LB/antibiotic mixtures. The bacteria were cultured at 37ºC until an O.D. of 0.6 at λ = 600 nm was achieved. The temperature was lowered to 20ºC, gene expression was induced with 1 mM IPTG and cultures were incubated overnight. Cells were then harvested by centrifugation (4,000 *g* at 4ºC for 10 minutes). Cultures were resuspended in a lysis buffer (buffer A, see Table 3.1) with the addition of complete EDTA-free protease inhibitor cocktail tablets from Roche. The cells were disrupted using a French press/cell disruptor, and homogenates were centrifuged at 40,000 *g* for 30 minutes at 4ºC. The resulting supernatants were passed through a 0.45 μm filter. Proteins were loaded on a 5-ml HisTrap HP column (GE Healthcare) to perform affinity chromatography (Ni-NTA). After a 10-column volume (CV) washing step using 90% buffers A and 10% of buffer B (buffer A and 0.8 M imidazole), the recombinant proteins were eluted applying a linear imidazole gradient. The eluted proteins were dialysed into buffer A and incubated overnight with 1 mg of His-tagged TEV protease per 20 mg of protein at 4ºC, then applied to a HisTrap HP column equilibrated with buffer A to remove the TEV protease, cleaved peptide and non-cleaved material. Fractions containing the proteins were collected, concentrated and applied to a size-exclusion chromatography (SEC) column (HR 16/60, Superdex 75 prep grade, GE Healthcare) equilibrated with buffer A. The SEC columns had been calibrated with molecular mass standards (thyroglobulin, 670 kDa; γ -globulin, 158 kDa; serum albumin, 67 kDa; ovalbumin, 44 kDa; myoglobin, 17 kDa; vitamin B_{12}, 1 kDa). Protein purity and molecular weight were assessed by SDS-PAGE (Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis, Biorad stain-free precast gels) and MALDI-TOF-MS (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry) analysis performed at the University of Dundee 'Fingerprints' Proteomics Facility using an Applied Biosystems Voyager DE-STR spectrometer. The stain-free SDS-PAGE gels do not require Coomassie Blue stain for protein

visualization. Instead, a trihalo compound covalently binds to tryptophan residues under UV irradiation resulting in UV-induced fluorescence (Short and Posch, 2011). Protein quaternary structure was investigated by native-PAGE (Invitrogen precast gels) and SEC. Specific details for the protein production of the three proteins are shown in Table 3.1.

Protein (MW, kDa)	Restriction enzymes (Insert size, bp)	Expression	Lysis buffer, Ni-NTA (buffer A)	Ni-NTA (buffer B)	SEC (buffer C)
*Pa*KynB (23.2)	NdeI/XhoI (645)	BL21(DE3) pLysS, 1mM IPTG, 16ºC O/N	25 mM Tris-HCl, pH 7.5, and 100 mM NaCl	Buffer A and 0.5 M imidazole	10 mM NaH_2PO_4/Na_2HPO_4 pH 7.8, 20 mM NaCl and 0.5 mM Tris-HCl
*Bc*KynB (23.2)	NdeI/XhoI (645)	BL21(DE3) pLysS, 1 mM IPTG, 16ºC O/N	25 mM Tris-HCl, pH 7.5, and 100 mM NaCl	Buffer A and 0.5 M imidazole	10 mM NaH_2PO_4/Na_2HPO_4 pH 7.8, 20 mM NaCl and 0.5 mM Tris-HCl
*Ba*KynB (22.7)	NdeI/BamHI (633)	BL21(DE3), 1 mM IPTG, 20ºC O/N	20 mM Tris-HCl, pH 7.4, and 200 mM NaCl	Buffer A and 0.8 M imidazole	Buffer A

Table 3.1: Protein production details for KynB including: the restriction enzymes used for the cloning, the protein expression conditions (*E. coli* strain, IPTG concentration and overnight (O/N) culture temperature), and buffers used during purification (buffers A, B and C). Protein molecular weight and corresponding gene size are annotated in brackets in the first two columns.

Protein concentrations were measured by absorbance at 280 nm using the Lambert-Beer law and the predicted molar extinction coefficients estimated using the ExPASy bioinformatics resource portal PROTPARAM (Gasteiger *et al.,* 2003): ε(*Pa*KynB)=28210 M^{-1} cm^{-1}, ε(*Bc*KynB)=28210 M^{-1} cm^{-1} and ε(*Ba*KynB)=20970 M^{-1} cm^{-1}.

3.2.2 Crystallographic analysis

3.2.2.1 Crystallisation and diffraction measurements

Pure *Pa*KynB and *Bc*KynB were concentrated to 7.5 mg ml^{-1} and *Ba*KynB to 4 mg ml^{-1} in buffer A to provide a stock solution for crystallisation experiments. Commercial

screens from Molecular Dimensions and Qiagen were used as the first crystallisation trials. These were performed in 96-well sitting drop plates with a Phoenix liquid handling system (Rigaku-MSC) using a ratio of 1:1 for protein solution:reservoir and final volumes of 0.2 and 0.4 µl for every condition. Then, plates were incubated at room temperature in a Gallery DT plate hotel (Rigaku-MSC).

In the case of *Pa*KynB, a number of promising conditions were identified from the crystallisation screens Morpheus (Molecular Dimensions), The Classics Suite (Qiagen), JSCG Plus (Molecular Dimensions) and MIDAS (Molecular Dimensions). Crystals were obtained after 2 to 30 days of incubation at 21ºC. Optimisation in hanging drop plates gave some improvement, however most of the crystals obtained were small (maximum dimension less than 15 µm) and the biggest *Pa*KynB crystal was around 40 x 20 x 20 µm in a condition containing 0.1 M HEPES pH 7.5, 0.1 M NaCl and 1.6 M $(NH_4)_2SO_4$. Crystals were tested in-house and at the Diamond synchrotron using from 20 to 80% (v/v) glycerol for cryo-protection and then flash frozen in liquid nitrogen. These crystals showed no diffraction. Parallel experiments carried out by Dr. V. Srikannathasan, gave improved *Pa*KynB crystals and the structure was determined. The reservoir condition contained 0.1 M Hepes, pH 7.5, 20% (w/v) PEG 4000 and 10% (v/v) propan-2-ol. The *Bc*KynB crystal was obtained by Dr. V. Srikannathasan using a reservoir comprising 13% (w/v) PEG 3350, 5 mM $CoCl_2$, 5 mM $CdCl_2$, 5 mM $MgCl_2$ and 5 mM $NiCl_2$.

Before data collection, crystals were incubated for a few seconds in a solution containing the reservoir condition and 25% (v/v) glycerol. Diffraction data for *Pa*KynB and *Bc*KynB were collected on the Diamond Light Source beamline I03 using a Pilatus 6M-F detector, indexed and integrated using XDS (Kabsch, 2010), scaled and analysed

with SCALA and POINTLESS (Evans, 2005) from the CCP4i suite (Winn *et al.,* 2011). The first *Ba*KynB crystals were obtained using The PEGs Suite (Qiagen) and The Classics Suite. Optimisation was carried out and small diffracting crystals obtained using a reservoir containing 0.1 M Tris-HCl pH 8.5, 0.15 M $MgCl_2$, and 30% PEG 4000. They were flash frozen in liquid nitrogen and tested on the in-house X-ray generator and Diamond Light Source I03. These samples were mechanically twinned and further optimisation was required. The experiment consisted of 24-well hanging drop plates, including different dioxane concentrations (from 1% (v/v) to 5% (v/v)), and a ratio of 1:1 in final volumes of 2 and 4 µl at 20°C. The presence of 1.5-2% (v/v) dioxane improved the quality of the sample and *Ba*KynB crystals were flash frozen directly from the drops in which they grew. Diffraction data were obtained using the in-house X-ray facility (Rigaku M007HF X-ray generator with a Saturn 944HG+ CCD detector). Data were indexed and integrated using iMOSFLM (Leslie and Powell, 2007) and scaled and analysed by AIMLESS (Evans, 2011) from the CCP4i suite. To obtain information about the binding of the product, co-crystallisation in the presence of L-kynurenine was carried out. If the substrate NFK were added, the reaction would have progressed whereas KynB requires the presence of both products, L-kynurenine and formate, for the reverse catalysis. Prior to crystallisation, the *Ba*KynB protein stock solution was incubated with 8 mM L-kynurenine for 10 minutes at room temperature. A compound that mimics the substrate and product structures, 2-aminoacetophenone (Figure 3.1), was also used for co-crystallisation. To obtain crystals of the *Ba*KynB:2-aminoacetophenone complex, the protein was incubated with 5% (v/v) of the compound before crystallisation. Preparation of the ligand stock at 20% (v/v) concentration was performed in a solution containing buffer A and 20% (w/v) PEG 4000. Diffraction data for the *Ba*KynB:2-aminoacetophene complex were obtained at the Diamond Light Source using beamline I03. The data showed the *Ba*KynB:2-aminoacetophenone crystal

is isomorphous with the previous *Ba*KynB crystals.

3.2.2.2 Structure determination

The first KynB structure, *Pa*KynB, was determined by Dr. V. Srikannathasan using molecular replacement (MOLREP, Vagin and Teplyakov, 2009) with one subunit of a putative metal-dependent hydrolase from *Geobacillus stearothermophilus* (*Gs*Hydrolase, PDB code 1R61). All water molecules and ions were removed from the search model. The *Gs*Hydrolase structure shares 25% sequence identity with *Pa*KynB. The resolution of the search model is 2.5 Å and one Zn^{2+} was included in the active site. However, the ion was assigned an occupancy of 0.5 and had a *B*-factor of about 90 $Å^2$. No publication is available so there are no details about the Zn^{2+} assignment. *Pa*KynB and *Bc*KynB share a sequence identity of 64% so the *Bc*KynB structure was solved using a partially refined *Pa*KynB as the search model (by Dr. V. Srikannathasan) and then the *Ba*KynB structure was solved using *Bc*KynB (40% sequence identity) as the search model. Refinement protocols for the different models were similar. These involved first rigid body refinement as part of the molecular replacement calculations, then iterative cycles of restrained refinement combining REFMAC (Murshudov *et al.,* 2011) with electron and difference density map inspections, and model manipulations in COOT (Emsley *et al.,* 2010). The starting *B*-factors for each model were derived from the Wilson *B*-factor. The number of peptide chains in the asymmetric unit was estimated using the Matthews coefficient (V_m), which refers to the crystal volume per unit of molecular weight. It is calculated from the volume of the unit cell, protein molecular weight, the number of asymmetric units within the unit cell and the number of molecules in the asymmetric unit (N). V_m values are calculated for different N and compared with the experimental data from the PDB. The probability of presenting a particular N is determined by the proximity of V_m to an experimental range

(Kantardjieff and Rupp, 2003). Determination of V_m and the percentage of bulk solvent (disorder solvent between protein molecules in the crystals, Weichenberger *et al*., 2015) values were carried out with CCP4i. Two polypeptide chains constitute the asymmetric unit of *Pa*KynB (showing a V_m of 3.6 $Å^3$ Da^{-1} and bulk solvent 66%) and four in the case of *Bc*KynB (V_m 2.8 $Å^3$ Da^{-1}, bulk solvent 56%) and *Ba*KynB (V_m 2.2 $Å^3$ Da^{-1}, bulk solvent 44%). Tight non-crystallographic symmetry (NCS) restraints were imposed at the beginning of all refinements, which were gradually released during the process. Additionally, the weighting of the external restrains was modified from the default parameters. The weighting values were obtained from PDB_REDO (Joosten *et al.,* 2014). Although L-kynurenine was present in the crystallisation mixture, no evidence of its presence was found in electron density maps. Even with a higher concentration of 2-aminoacetophenone present, this ligand was only observed in one active site. Once the protein structures were complete, alternative side-chain rotamers, water molecules, metal ions and ligands were incorporated into the models. A correction for a twinning component of 0.15, calculated in REFMAC, was applied in refinement of the *Ba*KynB:2-aminoacetophenone complex. Refinements were terminated when geometry of the model did not improve, there were no significant changes in R_{work} and R_{free} values and inspection of the difference density maps suggested that no further corrections or additions were justified. The assignment of the two cations present in the active sites is described below.

3.2.2.3 Cation identification

The difference Fourier maps suggested the presence of metals in the active site. The presence of Zn^{2+} was first confirmed in *Pa*KynB and *Ba*KynB crystals using X-ray absorption near-edge structure (XANES) spectroscopy measured at Diamond Light Source on beamline I03 using a Vortex Silicon Drift detector and an excitation energy

range from 9626.7 to 9690.14 eV. Since no Zn^{2+} was added during protein production, purification or crystallisation, then the most likely source is the media used to culture the *E. coli*. In the case of *Bc*KynB, crystals grew in the presence of different metal ions (5 mM $CoCl_2$, 5 mM $CdCl_2$, 5 mM $MgCl_2$ and 5 mM $NiCl_2$) so XANES spectroscopy was not suitable for the determination of the presence of non-added metal ions in the loop.

The electron density and anomalous scattering were used to identify the two cations present in the *Bc*KynB active site. Firstly, 12 sulfur atoms in the asymmetric unit, with an average *B*-factor of 16.4 ± 4.1 $Å^2$, were removed from the model to calculate their positive density in a difference Fourier map. The average peak height was 17.6 ± 3.2 σ what gives 1.1 σ per electron. In the case of the *Bc*KynB metal ion sites, the map showed positive density peaks with heights of 46.5 ± 4.6 σ (peak-1, corresponding with 42 e^-) and 68.4 ± 4.6 σ (peak-2, 62 e^-) in the four active sites. These height values indicate the presence of two different metal atoms per active site. The coordinating atoms surrounding the metal ions sites are more ordered than the 12 sulfur atoms removed for this calculation. The average *B*-factor of these atoms is 10.6 ± 1.1 $Å^2$ for the peak-1 site and 11.9 ± 2.4 $Å^2$ for peak-2 (1.5 and 1.3 times lower than sulfur, respectively). Taking this into account, a *B*-factor-adjusted calculation suggests that the ions sites are occupied by ions of approximately 28 e^- (peak-1) and 48 e^- (peak-2). Comparing these values with the atomic numbers of the metals present in the crystallisation condition suggested peak-1 corresponds with a Zn^{2+} ion and peak-2 with a Cd^{2+} since the expected values for these ions are 28 and 46 e^-, respectively. The average refined *B*-factors for Zn^{2+} and Cd^{2+} are 10.1 ± 1.1 and 11.4 ± 1.1 $Å^2$. Additional confirmation was obtained by an anomalous dispersion difference Fourier calculation. Data collection was performed at λ = 0.9795 Å. At this wavelength the theoretical f''

values for Zn^{2+} and Cd^{2+} are 2.480 and 2.132 e^- (Sasaki, 1989). Therefore, an anomalous difference Fourier would show a slightly larger signal at the Zn^{2+} position. The program FFT (Fast Fourier Transform) from the CCP4i suite was used to determine the anomalous difference Fourier map. The average heights of the peaks are 24.2 ± 1.6 σ at the Zn^{2+} site and 22.5 ± 1.8 σ at Cd^{2+} over the four subunits, and the difference between the heights was consistent in the four active sites of the asymmetric unit (1.7 ± 0.2 σ). This matches with the expected result confirming the presence of one Zn^{2+} and one Cd^{2+} at each active site.

3.2.2.4. Purification and structure determination of BaKynB in the presence of EDTA

Crystallographic data indicated two Zn^{2+} ions were present in the active site of *Ba*KynB. To check the involvement of the Zn^{2+} in the reaction, enzymatic assays were carried out in the presence of EDTA. This compound is a standard high affinity cation-chelating agent. No decrease in enzyme activity was observed when the assay buffer contained 100 mM EDTA (See Results and discussion, section 3.3.4). To check whether the Zn^{2+} ions can be removed or not by the EDTA, *Ba*KynB was incubated with 10 mM EDTA (*Ba*KynB-EDTA) at 4ºC overnight. As a control, *Ba*KynB in the absence of EDTA was also incubated under the same conditions. Excess EDTA was removed by SEC. These samples were assayed at 1 mM NFK concentration, in the presence and absence of 50 mM EDTA or 0.2 mM $ZnCl_2$.

*Ba*KynB-EDTA was crystallised, in the same conditions as *Ba*KynB, in the absence and presence of 5 mM $CdCl_2$ (*Ba*KynB-EDTA-Cd^{2+}) to test Cd^{2+} binding. Diffraction data were collected at Diamond Light Source beamline I03. Data reduction and analysis was performed using iMOSFLM and AIMLESS. The cell content analysis using the Matthew's coefficient suggested the presence of two peptide chains in the asymmetric

unit (V_m value of 2.1 $Å^3$ Da^{-1} and bulk solvent of 41% for both samples), indicating the crystals were not isomorphous with the previous *Ba*KynB (space groups and cell dimensions are detailed in section 3.3.2). The structures were solved using MOLREP and the *Ba*KynB dimer as the search model. A round of rigid body refinement using REFMAC was carried out and the difference Fourier map obtained. Peaks of positive density were observed in the Zn^{2+} binding sites (Table 3.2). Only two sulfur atoms were available in the structures (Met47) and used for the number of e^- determination for each peak. The results are similar to the other KynB. In the case of *Ba*KynB-EDTA structure, two Zn^{2+} are present in the active site. For *Ba*KynB-EDTA-Cd^{2+} the calculated e^- values are bigger than expected but the difference in the peak height indicates two different metal ions are present, as shown for *Bc*KynB. The crystallisation condition contained only two metals ions; Zn^{2+} and Cd^{2+}, so the cations were included in the model. For both structures, a round of refinement was performed to generate difference Fourier maps and *B*-factors of the metal sites. No difference map peaks at the cation positions, were observed. Reasonable *B*-factors were obtained in *Ba*KynB-EDTA-Cd^{2+}. In the case of *Ba*KynB-EDTA, higher *B*-factors were noticed (Table 3.2). This could be explained if not all the metal sites in the crystal were fully occupied. Full refinement of these structures was not continued. The presence of Cd^{2+} is consistent with *Bc*KynB results.

Protein	Peak height average, and calculated e^- number			*B*-factors average ($Å^2$)		
	Site 1	Site 2	Sulfur	Site 1	Site 2	Sulfur
*Ba*KynB-EDTA	12.5 ± 0.3 σ (25 e^-)	13.6 ± 1.1 σ (27 e^-)	7.8 ± 0.7 σ	46.3 ± 4.0	47.6 ± 1.3	37.9 ± 5.8
*Ba*KynB-EDTA-Cd^{2+}	15.0 ± 0.8 σ (38 e^-)	20.5 ± 2.1 σ (52 e^-)	6.3 ± 1.0 σ	19.0 ± 4.0	27.4 ± 5.7	28.4 ± 5.6

Table 3.2: Analysis of the metal ion sites in *Ba*KynB-EDTA and *Ba*KynB-EDTA-Cd^{2+} structures. Average of the difference Fourier map peaks, calculated e^- numbers and *B*-factors for metal sites and sulfurs.

3.2.2.5 Analysis of models and structure comparisons

MolProbity (Chen *et al.,* 2010) was used to assess geometry of all models. The *B*-factor average was calculated using the software BAVERAGE from the CCP4i suite. Secondary structure determination used both DSSP (Touw *et al.,* 2015) and visual inspection, and subunit interface surface area calculations were carried out with PISA (Krissinel and Henrick, 2007). Figures were prepared using ALINE (Bond and Schüttelkopf, 2009) and PyMOL (www.pymol.org). The DALI server (Holm and Rosenstrom, 2010) was used to search the PDB for structural homologues, whereas superpositions were calculated using DALILITE (Hasegawa and Holm, 2009).

3.2.3 Enzymatic assay

Kynurenine formamidase activity was measured using a spectrophotometric assay. The assay was optimised to obtain K_m, V_{max} and k_{cat} values for KynB. The assay is based on the increase in absorbance due to L-kynurenine formation (Katz *et al.,* 1987; Bougie, 2011), which is detected by measuring the absorbance at 365 nm with a UV-2450 Shimadzu spectrophotometer over a period of 160 seconds. Reactions were performed in triplicate using 1 ml final volume, at 25 ºC, using 0.025, 0.15, 0.25, 0.5, 0.7, 1, 1.5, 2 and 3 mM NFK substrate (Dalton Pharma Services), and 500 ng of enzyme (20 nM). The buffer used contained 0.1 M NaH_2PO_4/Na_2HPO_4 pH 7.4, and 20 µM $ZnCl_2$. His_6-tagged and non-His_6-tagged proteins were tested to compare activities. Also, two additional pH values were used. In these cases, the assay buffers consisted of NaH_2PO_4/Na_2HPO_4 pH 8.4 and pH 6.4. The initial velocity for each substrate concentration were calculated by fitting a linear equation to the data of the increase in L-kynurenine concentration during 1 min, and data were analysed using the Michaelis-Menten equation (SigmaPlot, Systat Software). The L-kynurenine concentration was calculated using the Lambert-Beer law and ε_{365} of 4220 M^{-1} cm^{-1} (Bougie, 2011).

3.2.4 Tryptophan fluorescence: fluorescence spectroscopy

Tryptophan fluorescence was used to investigate ligand binding. If the association of a ligand causes a conformational or environmental change in tryptophan residues, a difference in fluorescence may be observed. Fluorescence measurements for *Ba*KynB were performed using a LS-55 PerkinElmer spectrometer. To determine the maximum fluorescence due to tryptophan the excitation used was 280 nm and the emission wavelength detection ranged from 300 to 400 nm. The sensitivity of the detector was fixed to 800 V, and the peak of maximum emission was experimentally determined to be 337 nm. The sample (2 ml) contained 20 μg of *Ba*KynB (400 nM), 0.02 M Tris/HCl, pH 7.4 and 0.2 M NaCl. The two products of the KynB catalysed reaction, L-kynurenine and formate, and 2-aminoacetophenone were investigated separately. Concentrations ranged from 0 to 800 μM, with increments of 5 μM (for low concentrations), 10 μM (from 200 μM to 500 μM of ligand) and 100 μM (high concentrations) in the case of L- kynurenine and 2-aminoacetophenone. For formate, concentrations from 0 to 500 μM (same increments) were tested. A control experiment was carried out to check if there was fluorescence emission from L-kynurnine or 2-aminoacetophenone. The buffer in the presence of the maximum concentration of ligand and no protein present showed no fluorescence emission from either L-kynurenine or 2-aminoacetophenone under the experimental conditions. Results were represented as percentage of protein saturation within different ligand concentrations to determine K_d values using the one-site saturation binding curve using GraphPad Prism (www.graphpad.com).

3.3 Results and discussion

3.3.1 Recombinant protein production

Recombinant protein expression optimisation provided a yield of 10 mg L^{-1} in the case of *Pa*KynB and *Bc*KynB, and 53 mg L^{-1} for *Ba*KynB. SDS-PAGE indicated a high level of purity for the three samples. A native-PAGE and SEC were performed to investigate the *Ba*KynB quaternary structure. Results showed a band of approximately 43 kDa and a peak of about 46 kDa (SEC) indicating dimer formation. The MALDI-TOF trace also indicated highly pure samples had been obtained, and confirmed the monomer molecular weight (23.38 kDa). A peak corresponding with the dimer molecular weight (46.78 kDa) was observed (data not shown). Similarly, the MALDI-TOF trace of *Pa*KynB showed two peaks of 23.36 and 46.76 kDa.

In the case of *Pa*KynB, His_6-tag cleavage by TEV protease was not efficient and severe protein aggregation/precipitation was observed. To improve His_6-tag cleavage, different incubation temperatures and times were tested. Severe protein aggregation was observed when using temperatures lower than 30ºC, and incubation times longer than 2 hours at 30ºC. The enzyme assay (see later) indicated that the His_6-tag *Pa*KynB possessed a level of activity comparable with that of other samples, and the non-His_6-tag protein was inactive. Consequently, the His_6-tag of *Pa*KynB was left on for further experiments. Non-His_6-tagged *Ba*kynB and *Bc*KynB samples were always used unless otherwise indicated.

3.3.2 Crystallisation and structure determination

*Pa*KynB crystals were obtained for a range of crystallographic conditions, however they did not diffract (Figure 3.3). Eventually, diffracting crystals for *Pa*KynB and *Bc*KynB were obtained by Dr.V. Srikannathasan.

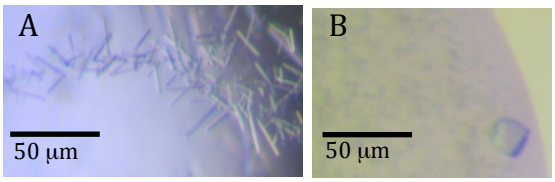


Figure 3.3: *Pa*KynB crystals. (A) Hanging drop: 0.05 M HEPES pH 7.5, 1% (v/v) PEG 400, 1 M $(NH_4)_2SO_4$ and 4 mg ml^{-1} protein in buffer A. (B) Sitting drop: 0.05 M HEPES pH 7.5, 0.05 M NaCl, 0.8 M $(NH_4)_2SO_4$, and 4 mg ml^{-1} protein in buffer A.

*Ba*KynB aggregated at relatively low concentrations (7 mg ml^{-1}). The protein stock was finally set up to 4 mg ml^{-1} in buffer A. Optimisation was carried out and resulted in crystals large enough to test on the in-house X-ray generator and at the Diamond Light Source beamline I03. Good diffraction was obtained from the crystal in Figure 3.4A but it showed twinning and the space group could not be assigned reliably. The crystal appearance and the splitting of the diffraction intensities (Figure 3.4B) suggest a mechanical twin. Further crystallisation optimisation was performed in the presence of various dioxane concentrations. Well-ordered diffracting crystals were obtained and structure determination pursued. Additionally, well-diffracting crystals grew in the presence of 2-aminoacetophenone and provided a dataset for determining a ligand:protein complex

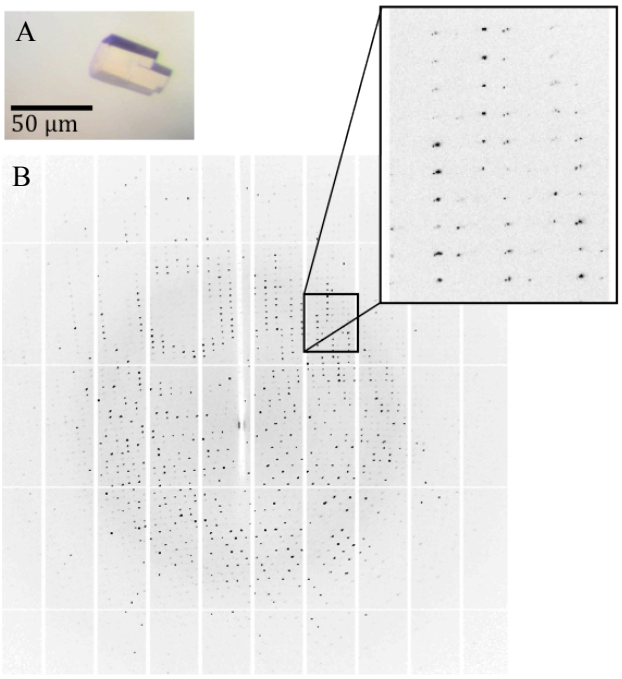


Figure 3.4: *Ba*KynB diffracting crystals. (A) Hanging drop: 0.05 M Tris-HCl pH 8.0, 0.15 M $MgCl_2$ and 15% (w/v) PEG 4000, and 2 mg ml^{-1} protein in buffer A. (B) Diffraction pattern from crystal A and an image expansion to show a mechanical defect.

structure. The crystallographic statistics for KynB are shown in Table 3.3. Statistics are shown in Table 3.4 for the crystals obtained after *Ba*KynB incubation with EDTA.

Structure	***Pa*KynB**	***Bc*KynB**	***Ba*KynB**	***Ba*KynB:ligand**
PDB code	4COB	4COG	4CO9	4CZ1
Space group	$P3_12_1$	$P2_1$	$P2_1$	$P2_1$
Wavelength (Å)	0.9795	0.9795	1.5418	0.9791
Unit cell dimensions *a*, *b*, *c* (Å)	112.7, 112.7 90.76	76.9, 50.1, 35.2, β=94.14º	73.2, 66.0, 83.8, β=90.32º	73.7, 66.6, 84.0, β=90.24º
Resolution range[a] (Å)	28.90 - 2.37	28.37 - 1.60	42.48 - 1.95	42.64 - 2.25
No. reflections	133546	479924	202165	104599
Unique reflections	27365	134904	58336	38046
Completeness (%)	99.2 (94.5)	99.3 (99.8)	99.4 (94.1)	98.1 (97.7)
R_{merge}[b]	0.057 (0.571)	0.068 (0.477)	0.063 (0.166)	0.164 (0.551)
Redundancy	4.9	3.6	3.5	2.7
< I/σ(I)>	16.8 (2.4)	12.8 (3.0)	11.1 (4.3)	4.8 (2.2)
Wilson *B* (Å^2)	43.87	14.97	11.84	15.33
R_{work}[c]/R_{free}[d]	0.1519 / 0.1945	0.1489 / 0.1842	0.1715 / 0.2066	0.1837 / 0.2241
DPI[e] (Å)	0.198	0.068	0.137	0.086
Bond lengths r.m.s.d. (Å) / angles[f] (º)	0.018 / 1.890	0.025 / 2.429	0.008 / 1.322	0.017 / 1.684
Average *B*-factors (Å^2)	51.4	16.4	15.3	20.9
Protein residues	412	831	829	825
Water molecules	181	1068	906	793
Metal ions	4 Zn^{2+}	4 Zn^{2+}, 4 Cd^{2+}, 8 Mg^{2+}	8 Zn^{2+}, 5 Mg^{2+}	8 Zn^{2+}, 3 Mg^{2+}
Ligands	-	12 glycerol, 10 1,2-ethanediol, 1 polyethylene glycol	2 dioxane, 3 1,2-ethanediol	2-amino acetophenone
Ramachandran analyses				
Favoured regions (%)	96.1	96.1	97.2	96.5
Allowed regions (%)	100	100	100	100

Table 3.3: Crystallographic statistics. [a.] Values in parentheses refer to the highest resolution shell. [b.] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. [c.] $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. [d.] R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations. [e.] Diffraction Precision Index (Cruickshank, 1999). [f.](Engh and Huber, 1991).

Structure	***Ba*KynB-EDTA**	***Ba*KynB-EDTA-Cd^{2+}**
Space group	$P2_1$	$P2_1$
Wavelength (Å)	0.9791	0.9791
Unit cell dimensions *a*, *b*, *c* (Å)	65.8, 44.8, 67.8 β=107.6	65.9, 44.9, 68.2 β=108.0
Resolution range[a] (Å)	53.92 - 2.30	64.88 - 2.7
No. reflections	60829	60831
Unique reflections	16737	10690
Completeness (%)	98.6 (98.2)	100 (100)
R_{merge}[b]	0.115 (0.651)	0.171 (0.638)
Redundancy	3.6 (3.5)	5.7 (6.0)
$< I/\sigma(I)>$	6.8 (2.4)	11.8 (7.2)
Wilson *B* ($Å^2$)	24.49	21.67
R_{work}[c]/R_{free}[d]	0.1919 / 0.2725 *	0.2103 / 0.2779 *

Table 3.4: Crystallographic statistics. [a.] Values in parentheses refer to the highest resolution shell. [b.] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - <I(hkl)>|/ \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $<I(hkl)>$ is the mean value of $I_i(hkl)$ for all *i* measurements. [c.] $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. [d.] R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations. (*) Refinement not completed.

3.3.3 KynB structure

The high structural homology between the KynB structures means that only one representative needs to be described in detail, concentrating on *Ba*KynB since there are high resolution apo and ligand-bound structures. Four subunits (A, B, C and D) are found in the asymmetric unit of *Ba*KynB and *Bc*KynB, and two (A and B) in *Pa*KynB.

3.3.3.1 Overall structure and dimer formation

KynB structure shows a new amidase fold unrelated to KFase. Each monomer contains 12 β-strands, four α-helices and three short segments of 3_{10}-helix between β4 and β5 (Figure 3.5). Eight of the β-strands form a β-barrel-like domain if considering β1 and β4 as a single strand. One side of the β-barrel-like domain is surrounded by three α-helices. This fold is highly conserved within the KynB structures, consistent with the high degree of their amino acid sequence identity. KynB homologues for Gram-negative (*Bc*KynB and *Pa*KynB) share about 64% identity while Gram-positive (*Ba*KynB) and Gram-negative have 40% (Figure 3.6).

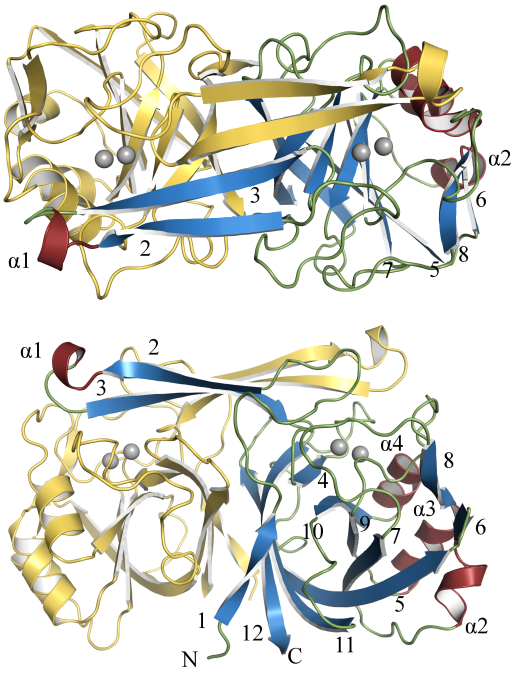


Figure 3.5: Cartoon representation of *Ba*KynB dimer. The front (top image) and lateral (bottom) views are shown. Subunit A: blue (β-strands), red (α-helices) and green (loops). Subunit B: yellow. α-helices are numbered using α followed by numbers. β-strands are named with numbers. Zn^{2+} ions are represented as grey spheres.

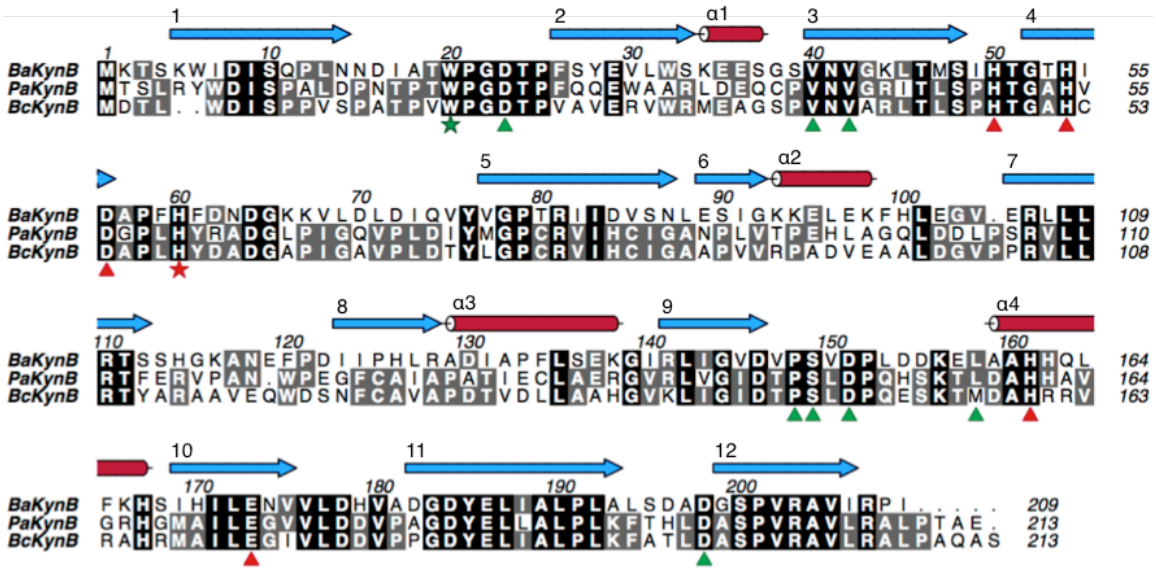


Figure 3.6: Alignment of *Ba*KynB, *Pa*KynB and *Bc*KynB amino acid sequences. Highly conserved residues are annotated in black, and conserved in two of the proteins in grey. *Ba*KynB secondary structure is represented as red cylinders (α-helices) and blue arrows (β-strands). Amino acids present in the active site pocket are indicated with green triangles. Red triangles and red star mark residues implicated in enzyme catalysis. The green star indicates a conserved tryptophan in the active site pocket.

The KynB crystal structures showed dimer formation (A-B and C-D subunits, Figure 3.5). This is consistent with the results from native-PAGE and SEC analysis. The dimer interface is formed by interaction of about 25% of the accessible surface area of each subunit. It involves the interaction of β1, β11, β12, and short helical segments between residues 68 and 74 from both subunits. The dimer arrangement results in the formation of a characteristic feature of this fold, an antiparallel four-stranded β-sheet involving β2-β3 from each subunit. The arrangement of the subunits is also highly conserved. The overlay of the KynB monomers and dimers (Appendix B) shows a high structural homology between the different homologues. In *Ba*KynB and *Bc*KynB, subunits A and C have higher structural similarity than A and B (or D and C), and the same feature applies to B and D subunits. Dimer overlay (Figure 3.7) gives r.m.s.d. (root-mean-square deviation) values between the Gram-negative representatives of 0.7 Å, and between Gram-positive and Gram-negative about 1.25 Å (Appendix B). More

specifically, residues that are conserved in all KynB subunits are positioned in the active site pocket and the β-barrel like domain.

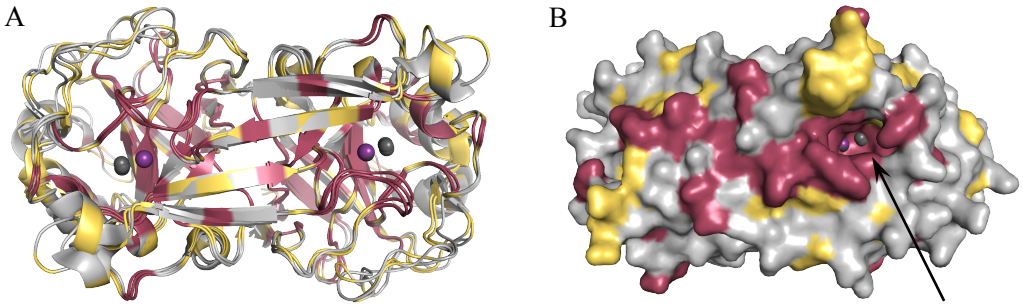


Figure 3.7: (A) Overlay of the C_α atoms from *Ba*KynB, *Pa*KynB and *Bc*KynB. Colours indicate the sequence identity; red (residues conserved in every protein), yellow (residues conserved in two of the proteins), grey (non-conserved residues). (B) Surface representation of the *Ba*KynB dimer and coloured with sequence identity conservation from A. The arrow indicates the active site pocket.

3.3.3.2 The KynB active site

The active site pocket is mainly formed by residues from loops between β1-β2, β3-β4 and β4-β5, from α4 and β10 from one subunit and β3 from the other (Figure 3.5). Structural comparisons indicate that the details in the active sites are conserved (Figure 3.8). In the active site two cations interact with His50, His54, Asp56, His161 and Glu173 in *Ba*KynB and *Bc*KynB, and His48, His52, Asp54, His160 and Glu172 for *Pa*KynB. There is also a residue that might be important for substrate recognition (see later), Trp20 (Trp18 in *Pa*KynB) that confers a hydrophobic environment. An investigation of available KynB amino acid sequences was conducted. First, a search of the bacterial kynurenine formamidase proteins annotated in UniProt (www.uniprot.org) was performed. This gave 380 unique sequences. The search was then filtered to 336 entries that share 30% sequence identity or more with *Ba*KynB using the NCBI BLAST server (National Center for Biotechnology Information Basic Local Alignment Search Tool, Johnson *et al*., 2008). Sequences were aligned using Clustal Omega (Sievers *et al.,* 2011). The result showed His54, Asp56 and His60 are strictly conserved in all

sequences, while His50, His161 and Glu173 are present in 99.7% of the entries. Trp20 is conserved in 81% of the entries and in 17.8% phenylalanine and tyrosine occupy the same position. These results confirm the importance of cation-interacting residues and the presence of an aromatic group in the active site pocket.

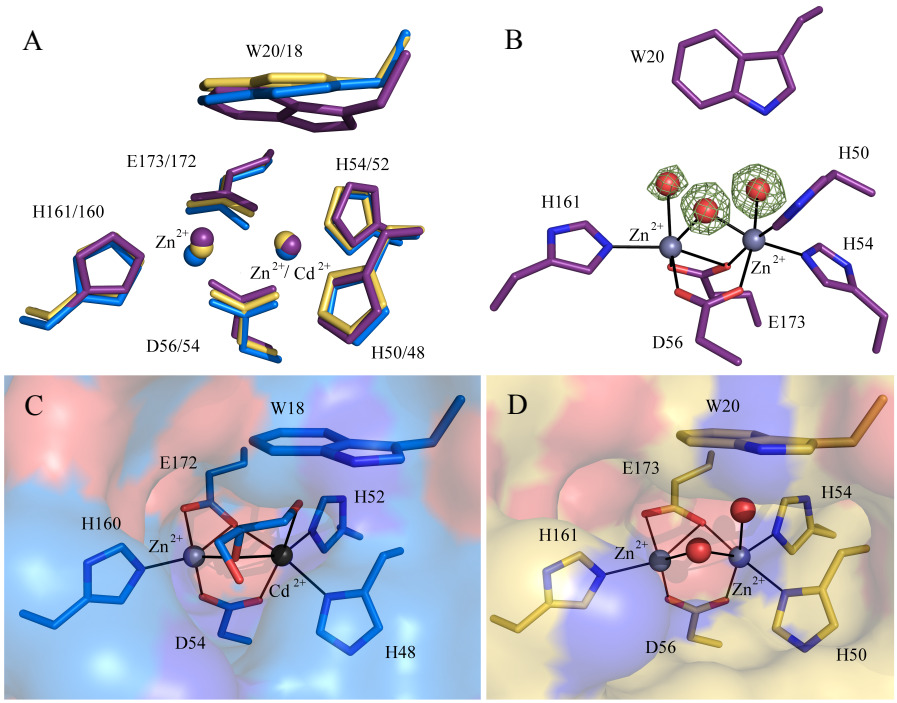


Figure 3.8: KynB active site. (A) Superposition of *Ba*KynB (violet), *Bc*KynB (blue) and *Pa*KynB (yellow) active sites. (B) The *Ba*KynB active site with difference density for the water molecules (red spheres) showed as green chicken wire and contoured at 5σ. (C) The *Bc*KynB active site with glycerol bound to the metal ions. (D) The *Pa*KynB active site. Coordination contacts are shown as continuous lines and Zn^{2+} ions as grey spheres.

Two Zn^{2+} ions were first found in the *Pa*KynB and *Ba*KynB active sites. In the case of *Bc*KynB one Zn^{2+} and one Cd^{2+} were present in each active site. The presence of the Zn^{2+} ions in *Pa*KynB and *Ba*KynB was confirmed by XANES measurements (Figure 3.9) and difference Fourier maps. *Bc*KynB cation characterisation was assessed by difference Fourier and anomalous difference Fourier maps (Figure 3.10). Three ordered waters were observed binding the cations in *Ba*KynB and generate a Zn^{2+} coordination number of six. For *Pa*KynB, only two waters are visible, probably due to lower resolution, and in *Bc*KynB waters have been replaced by glycerol (Figure 3.8).

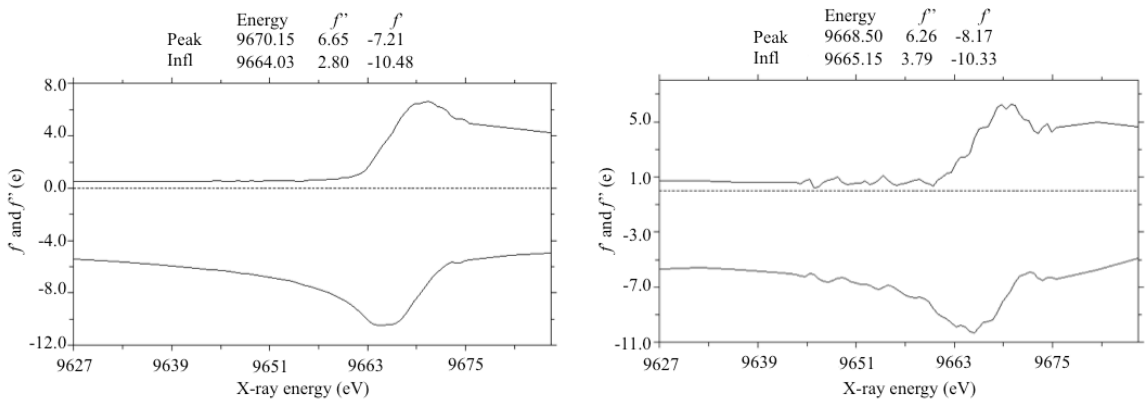


Figure 3.9: XANES spectra. K-edge XANES spectra showing the anomalous scattering factor *f'* and *f''* values within a range of X-ray energy from 9626.7 to 9690.14 eV. (A) *Ba*KynB (B) *Pa*KynB.

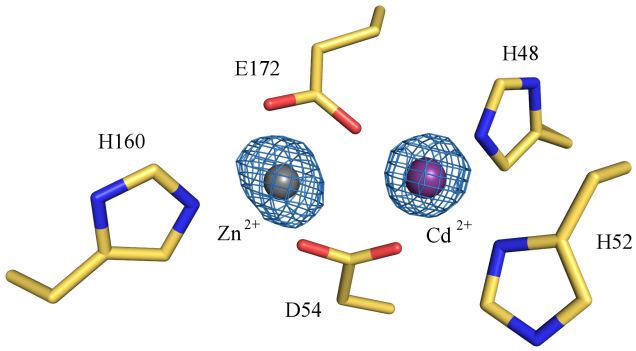


Figure 3.10: Anomalous difference Fourier map for *Bc*KynB. The map is shown as blue chicken wire (5.0 σ).

The structure of the *Ba*KynB:2-aminoacetophenone complex provides information about substrate recognition (Figure 3.11). Compound binding was first confirmed using tryptophan fluorescence (see section 3.3.5). In the structure, 2-aminoacetophenone is aligned parallel to Trp20. This model also provides information about the reaction mechanism. The amino group and the oxygen from the keto group form an intramolecular hydrogen bond that will stabilise the conformation of the intermediate at the active site. The occurrence of the hydrogen bond was confirmed by a search of structurally related compounds at the Cambridge Structural Database (CSD, Allen, 2002). A search of the two 2-aminoacetophonone conformations (with and without the intramolecular hydrogen bond) was carried out and identified 28 compounds containing the ligand scaffold. Inspection of these compounds showed the intramolecular hydrogen bond was present in all cases.

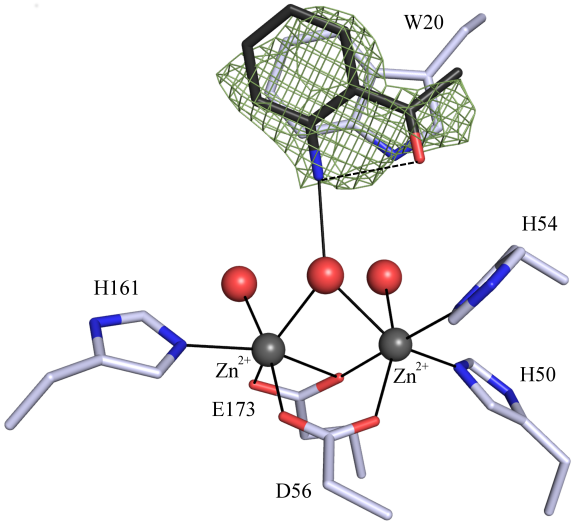


Figure 3.11: 2-aminoacetophenone (black) in the *Ba*KynB active site. The ligand intramolecular hydrogen bond is represented as a dashed line, and water molecules as red spheres.

A comparaison with available structures was performed using the DALI web server. Results revealed similarity between KynB and three proteins. These are a putative metal-dependent hydrolase/cyclase (PDB codes 3KRV and 1R61, sequence identity 25%) carrying one Zn^{2+}, an isatin hydrolase containing a single Mn^{2+} (Bjerregaard-Andersen *et al.,* 2014, sequence identity 23%) and a hypothetical protein with no metal ion present (2B0A, sequence identity 23%). Proteins with PDB codes 3KRV, 1R61 and 2B0A were solved by the Protein Structure Initiative (Gabanyi *et al.,* 2011) and there are no accompanying publications. The isatin hydrolase, an α/β hydrolase, shows similarities with KynB in the overall structure, the active site residues and aspects of the metal ion binding (Figure 3.12). It has one Mn^{2+} ion bound to an Asp, and two His that are positioned in a similar manner to His50, Asp56 and His60 in *Ba*KynB. Instead of a second metal, a water molecule is present contacting with the same Asp as Mg^{2+} and the other His (His161 in *Ba*KynB). The residues at the equivalent positions of Trp20 and Glu173 are not conserved. Indeed, the Glu173 is replaced by an alanine. This might prevent the incorporation of a second metal ion in the active site. A leucine residue is placed in the Trp20 position maintaining the hydrophobic properties of this site.

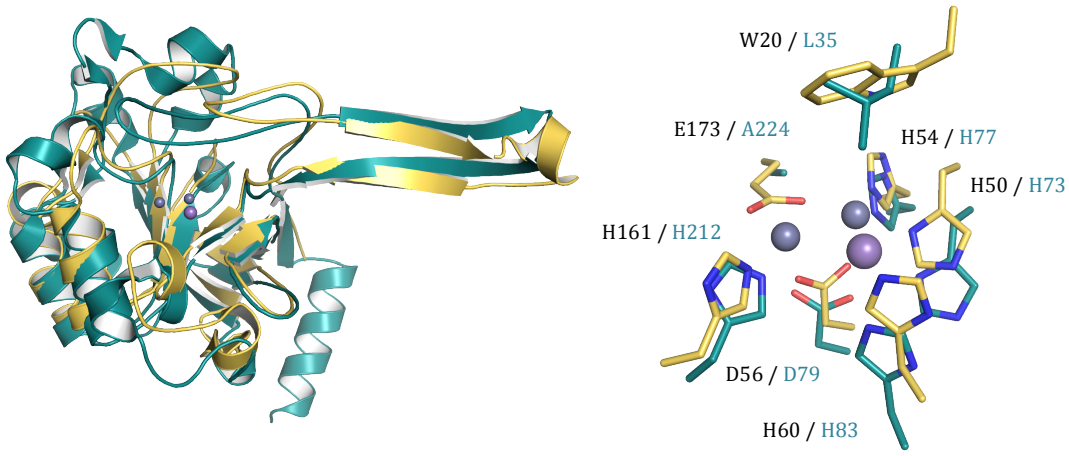


Figure 3.12: (A) Cartoon representation of the C_α overlay of *Ba*KynB (yellow) and isatin hydrolase (4J0N, green). Loops have been smoothed to facilitate the interpretation. (B) Sticks representation of the active sites pockets. Grey spheres: Zn^{2+} ions from *Ba*KynB. Purple sphere: isatin hydrolase Mn^{2+} ion.

3.3.3.3 Reaction mechanism

The presence of the two metal ions in KynB indicate a different mechanism to KFase and will likely follow acidic hydrolysis. The mechanism may be similar to that proposed for metalloenzymes from the amidohydrolase superfamily such as type B β-lactamases (Schenk *et al.,* 2012; Bebrone, 2007), phosphotriesterases (Bigley and Raushel, 2013), leucine aminopeptidases (Sträter *et al.,* 1999) and dihydro-orotases (Lee *et al.,* 2007). This family presents flexible loops at the active site important for substrate recognition (Seibert and Raushell, 2005). In KynB, the catalytic site is relatively rigid. More specifically, *Ba*KynB Zn^{2+}-binding residues have an average *B*-factor of 9.6 $Å^2$ while the overall average is 15.1 $Å^2$. KynB functions with a binuclear Zn^{2+} active center which shares similarity with other metalloenzymes, for example phosphodiesterases and dihydro-orotases. In these proteins, the two Zn^{2+} are 3.5-4.0 Å apart and a low coordination number (four) allows the formation of an intermediate state by increasing coordination to five or six (Holm and Sander, 1997; Seibert and Raushel, 2005). In KynB, the cations are 3.1 Å apart in a more crowded environment and already have a coordination number of six. This suggests the coordinating water might be important for catalysis.

Following these considerations, a mechanism of catalysis can be proposed (Figure 3.13). The binding of the substrate is driven by the presence of Trp20. The acidic environment provided by the binuclear Zn^{2+} decreases the water pK_a. Then His60 acquires a proton from the water and generates a hydroxide. This nucleophile attacks the carbonyl component of NFK. The intermediate formation will probably involve interaction with Ser149. The products are formed by proton donation from His60 to the amine and subsequently C-N bond breakage (Figure 3.13).

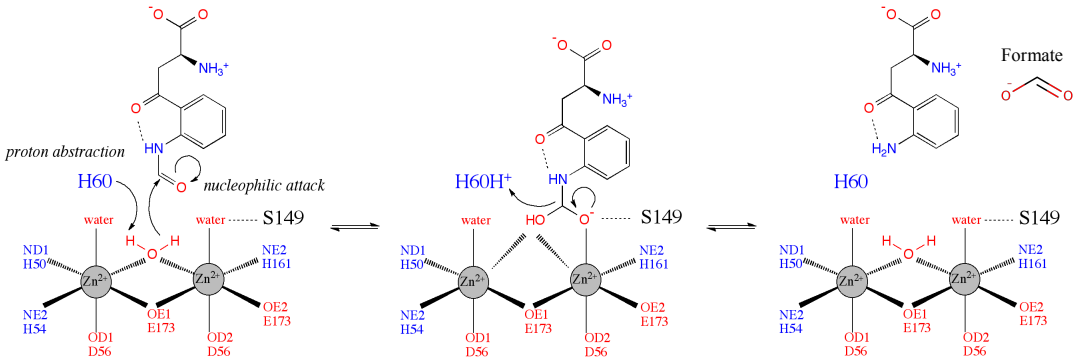


Figure 3.13: Proposed reaction mechanism.

3.3.4 Enzyme kinetics

Enzyme assays confirmed the KynB proteins presented kynurenine formamidase activity and allowed the determination of kinetic parameters for different protein samples and pH values (Table 3.5). The assay was performed with His_6-tagged and non-His_6-tagged proteins. In the case of *Pa*KynB, only the protein containing the His_6 tag was active but the other enzymes showed similar activities for both samples. Two additional pH values were tested for *Ba*KynB. At pH 6.4 the activity remained similar to pH 7.4 and pH 8.5 showed a slightly higher activity.

Protein	V_{max} (nmol min^{-1})	K_m (mM)	k_{cat} (s^{-1})	k_{cat} / K_m (M^{-1} s^{-1}) x10^4	SA (U mg^{-1})	EDTA effect
*Ba*KynB pH 7.4*	65.41 ± 2.64	0.40 ± 0.05	50.56	12.64	130.82	No
*Ba*KynB-His_6 pH 7.4	48.45	0.44	37.45	8.51	96.90	No
*Ba*KynB pH 6.4	81.27	0.67	62.82	9.38	162.54	No
*Ba*KynB pH 8.5	142.83	0.82	110.40	13.46	285.66	No
*Bc*KynB pH 7.4*	58.15 ± 0.98	0.57 ± 0.02	43.94	7.71	116.30	No
*Bc*KynB-His_6 pH 7.4	57.39	0.78	43.37	5.56	114.78	No
*Pa*KynB-His_6 pH 7.4*	147.99 ± 8.42	0.98 ± 0.13	114.21	11.65	295.98	No
*Bce*KynB	-	0.17	31.1	18.3	68.5	No
*Sp*KynB	-	0.31	-	-	26.8	No
*Rm*KynB	-	0.075	0.77	1.03	3.5	Yes

Table 3.5: Kinetic parameters of KynB samples. V_{max}: maximum catalytic velocity. K_m: Michaelis-Menten constant. Specific activity (SA): amount of active enzyme from the total protein quantity. k_{cat}: number of substrate molecules that are transformed per active site and per time unit. k_{cat}/ K_m: defines the catalytic efficiency. EDTA effect: if there was an effect in enzyme activity when EDTA was present in the reaction solution. *Bacillus cereus* (*Bce*KynB), *Streptomyces parvulus* (*Sp*KynB) and *Ralstonia metallidurans* (*Rm*KynB). (*) Samples used for crystallisation. Samples in blue indicate the assay was performed in triplicates and single assays are indicated in black.

The kinetic parameters for *Pa*KynB-His_6, *Ba*KynB and *Bc*KynB are similar to each other. *Streptomyces parvulus* and *Ralstonia metallidurans* (Katz *et al.,* 1987) KynB showed lower activities than the tested proteins. However, the reactions were carried out at 30°C and 37°C respectively, so the variation in activity might be due to the assay conditions. In the case of *Bacillus cereus* KynB (Bougie, 2011), the assays were performed at room temperature and the values are comparable to the data shown in Table 3.5.

During the assay, at tested conditions and NFK concentrations, no substrate inhibition was observed and protein saturation was obtained at around 2 mM of NFK. Activity for

*Ba*KynB-His$_6$ and *Bc*KynB-His$_6$ is similar to the non-tagged proteins, which were used for further assays. Higher activity is shown at pH 8.5 (Figure 3.14).

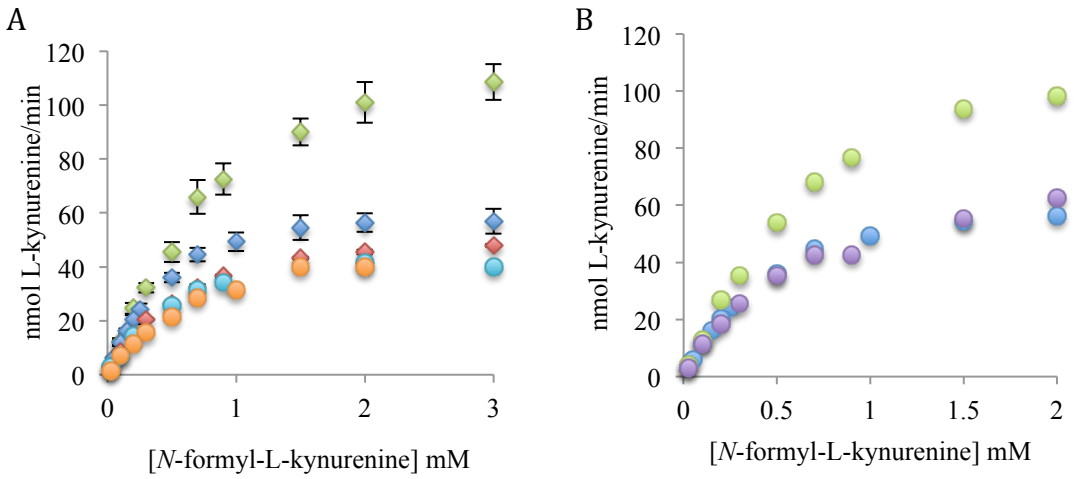


Figure 3.14: (A) Initial velocity plotted against substrate concentration for the different protein samples; (green) *Pa*KynB-His$_6$, (dark blue) *Ba*KynB, (red) *Bc*KynB, (light blue) *Ba*KynB-His$_6$, (orange) *Bc*KynB-His$_6$. (B) Initial velocity against substrate concentration at three different pH values; (green) *Ba*KynB pH 8.5, (purple) *Ba*KynB pH 6.4, (blue) *Ba*KynB pH 7.5.

*Ba*KynB and *Pa*KynB structures showed the presence of two Zn^{2+} ions at the active site. Assays were performed in the presence of different concentrations of EDTA. If the cations are required for catalysis, then the enzyme should have a lower enzyme activity in the presence of EDTA compared to the normal assay. Initial assays showed no activity decrease when EDTA was added up to 100 mM. Subsequently, the following experiment was performed.

*Ba*KynB was incubated with and without 10 mM EDTA (*Ba*KynB-EDTA) at 4ºC overnight before SEC. These samples were then assayed in the presence and absence of EDTA or $ZnCl_2$. Results showed a very small decrease in activity for *Ba*KynB-EDTA in comparison with the control sample (Table 3.6). Comparing the samples in the presence and absence of 50 mM EDTA, *Ba*KynB-EDTA samples showed 1.8 times less initial velocity. When 0.2 mM $ZnCl_2$ was added, *BaK*ynB-EDTA sample increased its

activity 2.5 times and *Ba*KynB sample 1.5 times. Assuming, Zn^{2+} is important for catalysis, incomplete removal of the Zn^{2+} ions by the EDTA could explain this result. The presence of a small amount of Zn^{2+}-free protein may explain the decrease in activity. The increase in protein activity by the addition of $ZnCl_2$ could be an indication for an activating role of Zn^{2+}. In addition, crystallography was performed with *Ba*KynB-EDTA to confirm if the Zn^{2+} was completely removed after the EDTA incubation. Diffracting crystals were obtained for *Ba*KynB-EDTA and *Ba*KynB-EDTA incubated with 5 mM $CdCl_2$ (*Ba*KynB-EDTA-Cd^{2+}). *Ba*KynB-EDTA revealed two Zn^{2+} ions were present at the active site. Data from *Ba*KynB-EDTA-Cd^{2+} showed one Zn^{2+} and one Cd^{2+} at the active site of every subunit (Figure 3.15).

Protein	EDTA 50 mM	0.2 mM $ZnCl_2$	V_o (nmol min^{-1})
*Ba*KynB-EDTA	-	-	15.6
*Ba*KynB-EDTA	+	-	17.1
*Ba*KynB-EDTA	-	+	38.4
*Ba*KynB	-	-	27.0
*Ba*KynB	+	-	29.9
*Ba*KynB	-	+	41.2

Table 3.6: Summary of the EDTA experiment results. *Ba*KynB-EDTA: sample incubated with 10 mM EDTA over night. *Ba*KynB: sample incubated without EDTA overnight. In column two it is indicated if 50 mM EDTA was added (+) or not (-) to the assay mixture. In column three it is indicated if 0.2 mM $ZnCl_2$ was added (+) or not (-) to the assay mixture. V_o: initial velocity in nmol of L-kynurenine per minute.

These results suggest that either Zn^{2+} has not been removed after protein incubation with EDTA or, taking into account the enzymatic data, a small amount of the cation was removed. One explanation might be that the active site is not accessible for EDTA to chelate and extract the cations. That Cd^{2+} can replace one of the Zn^{2+} ions is probably due to a higher affinity for one of metal sites, together with the high concentration used.

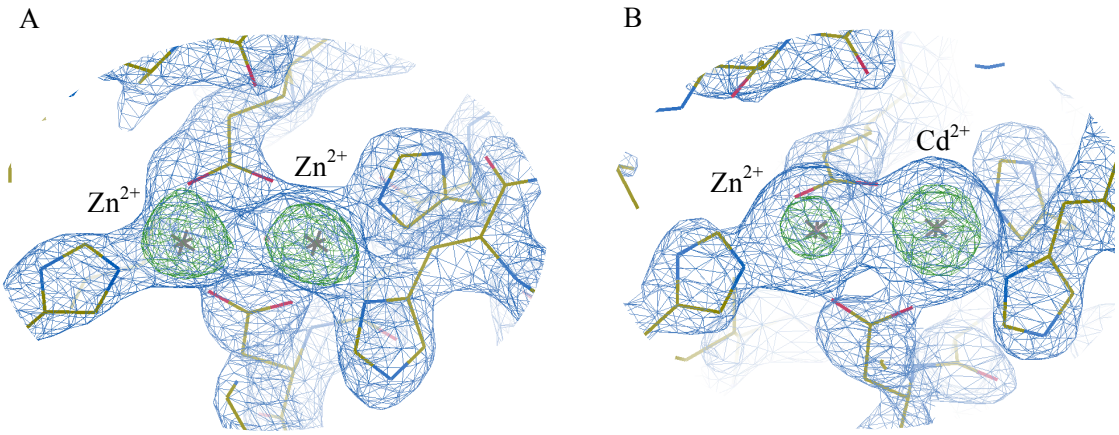


Figure 3.15: Difference Fourier peaks for metal sites. Blue maps are the electron density maps at contoured level 2.0 σ, and the difference maps are shown in green (A) *Ba*KynB-EDTA, difference map at 5.0 σ. (B) *Ba*KynB-EDTA-Cd^{2+}, difference map at 11.0 σ.

3.3.5 Fluorescence spectroscopy of *Ba*KynB

The presence of a tryptophan at the active site pocket allowed the study of binding of various compounds using fluorescence spectroscopy. Binding of L-kynurenine, formate and 2-aminoacetophenone with *Ba*KynB was investigated. When L-kynurenine and 2-aminoacetophenone were added there was a decrease of the fluorescence emission and shift of the maximum from 337 nm to 310 nm (Figure 3.16).

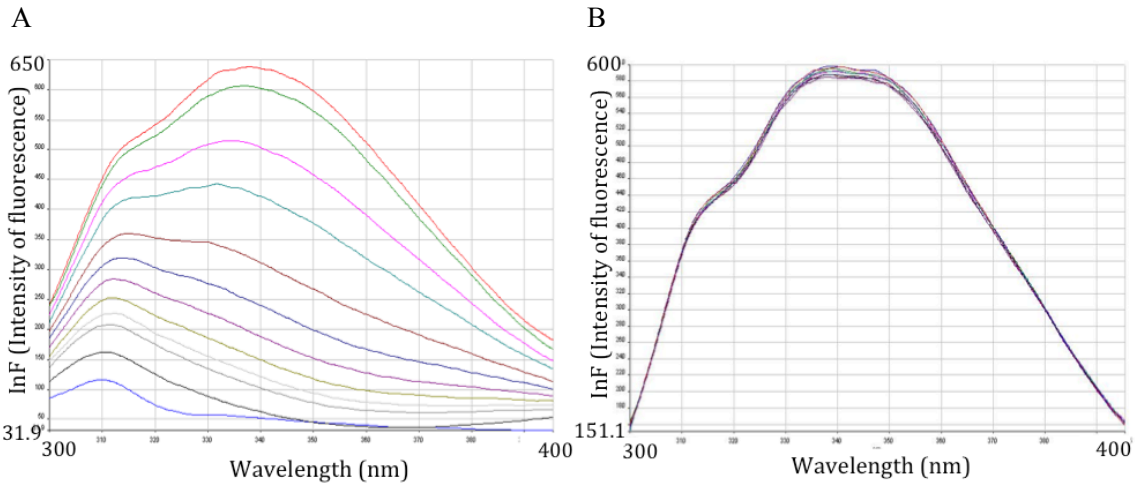


Figure 3.16: Emission spectra from *Ba*KynB using an excitation wavelength of 290 nm. (A) In the presence of L-kynurenine. Titration with 2-aminoacetophenone gave similar results. Red line: Without L-kynurenine. (Maximum InF peak at 337 nm). Top green line: Lowest L-kynurenine concentration tested. Dark blue bottom line: Highest concentration of L-kynurenine tested. (Shift of the maximum InF to 310 nm). (B) In the presence of formate. Maximum InF peak at 337 nm.

This confirmed the binding of 2-aminoacetopenone and L-kynurenine with K_d values of 60 µM and 50 µM, respectively (Figure 3.17). In the case of L-formate, no differences in fluorescence were detected with the increasing concentrations of the compound. These results are consistent with the proposed reaction mechanism. As inferred from *Ba*KynB:2-aminoacetophenone structure, L-kynurenine and 2-aminoacetophenone bind parallel to the tryptophan. This creates quenching of the fluorescence emission and the binding can be detected. However, the location of formate in the active site would be far from the tryptophan and no interference with the fluorescence will be observed.

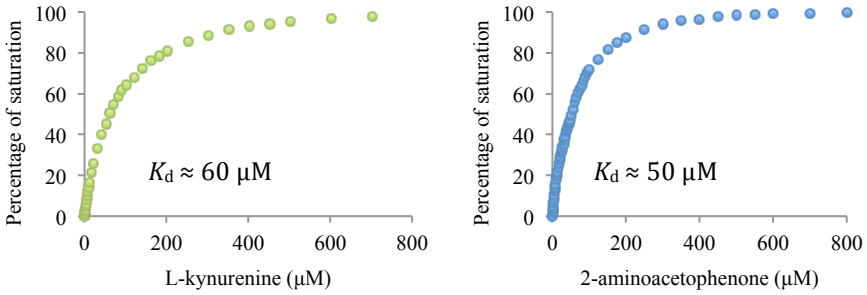


Figure 3.17: Plot of the percentage of enzyme saturation against L-kynurenine (green) and 2-aminoacetophenone (blue) concentration.

3.4 Conclusions

Three KynB enzymes from different bacteria, including Gram-negative and Gram-positive representatives, were cloned, expressed, purified, crystallised and the kynurenine formamidase activity confirmed. The structures showed a new amidase fold highly conserved across these species. The active site contains two Zn^{2+} ions in a rigid and crowded environment. The presence of Zn^{2+} indicates the amidase reaction takes place in an acidic environment and the mechanism diverges from KFase. The structure of a *Ba*KynB:2-aminoacetophenone complex gave information about substrate binding

suggesting Trp20 is a key feature in KynB. Moreover, it provides information to propose a distinct amidase reaction mechanism.

The differences found in protein structure and enzyme mechanism between the prokaryote kynurenine formamidase and the eukaryote KFase suggest KynB is a potential drug target. Moreover, optimisation of protein crystallography and enzyme assays made this target suitable for compound screenings and hit validation. However, results on essentiality from the Dstl revealed it is not an essential target in either *B. anthracis* or *B. pseudomallei*. For this reason, no further work was carried out.

PART FOUR

D-ALANINE—D-ALANINE LIGASE

4.1 Introduction

D-alanine (D-Ala) is a key amino acid in bacteria. It contributes to structural roles in peptidoglycan (PG) layer formation and in lantibiotics (Cava *et al.,* 2010). This D-amino acid is particularly important in Gram-positive bacteria where is linked to lipoteichoic acid (LA) by an ester bond (D-Ala:LA) (Childs and Neuhaus, 1980). The amount of D-Ala:LA in the PG layer regulates ion homeostasis, cell wall permeability, biofilm formation, cell adhesion and mediates drug resistance (Neuhaus and Baddiley, 2003). An increase in temperature, pH and NaCl/KCl concentration results in a decrease in D-Ala:LA formation. This leads to a change in the electromechanical properties of the cell wall, producing a more electronegative net charge of the layer. In terms of cation binding (*e.g.* Mg^{2+}, Ca^{2+}), a lower D-Ala ester level increases wall permeability to these components and serves to maintain cation homeostasis in the cell. The existence of D-Ala esters also prevents the binding of endolysins to LA influencing their effects. Regarding drug sensitivity, the decrease in D-Ala:LA enhances the antibacterial potency of cationic compounds (*e.g.* vancomycin, teicoplanin and balhimycin) and bactericidal enzymes. Human phospholipase A_2 acts as an anti-*Staphylococcus* agent. It is a Ca^{2+}-dependent enzyme, and a higher level of Ca^{2+} in the peptidoglycan layer results in a greater enzyme activity. In similar fashion, a higher amount of D-Ala:LA leads to a more positively charged layer and has been shown to be involved in antibacterial resistance to cationic compounds. LA has an important role in biofilm formation and cell adhesion. These processes are also Ca^{2+}-dependent and the lack of D-Ala:LA results in the lost of cell adhesion. In this case, the function of LA is

the presentation of Ca^{2+} to adhesins necessary for cell aggregation. This extends to virulence, where adhesion to the host tissue is also limited by D-Ala:LA levels (Neuhaus and Baddiley, 2003).

The alanine enantiomers can be transported into the cell by an Ala/Gly/Ser transporter, a H^+ symporter (Robbins and Oxender, 1973) or synthesised by bacteria. The route for the biosynthesis of D-Ala is through the enzyme alanine racemase (AlaR, EC 5.1.1.1, Zajdowicz *et al.,* 2011) and an important role of this D-amino acid is the formation of the PG layer precursor D-alanyl—D-alanine (Bugg *et al.,* 2011). The enzyme involved in the synthesis of the dipeptide is D-alanine—D-alanine ligase (Ddl, EC 6.3.2.4), the second enzyme of the PG biosynthesis pathway. D-Ala—D-Ala will link to UDP-*N*-acetyl-muramyl-L-Ala-D-Glu-*meso*-diaminopimelate (UDP-NAcMur-A-E-mDAP) to form a pentapeptide (UDP-NAcMur-A-E-mDAP-D-Ala—D-Ala). Then MraY (phospho-NAcMur-pentapeptide translocase, EC 2.7.8.13) removes UMP generating phospho-NAcMur-A-E-mDAP-D-Ala—D-Ala that binds to a membrane acceptor (undecaprenyl phosphate). Then, MurG (peptidoglycan glycosyltransferase, EC 2.4.1.227) incorporates *N*-acetyl-glucosamine (NAcGlc) in the phosphate position. It is transferred to the periplasmic space by flippases (still to be identified) where PBPs (penicillin-binding proteins, EC 3.4.16.4) crosslink and polymerise NAcGlc-NAcMur-

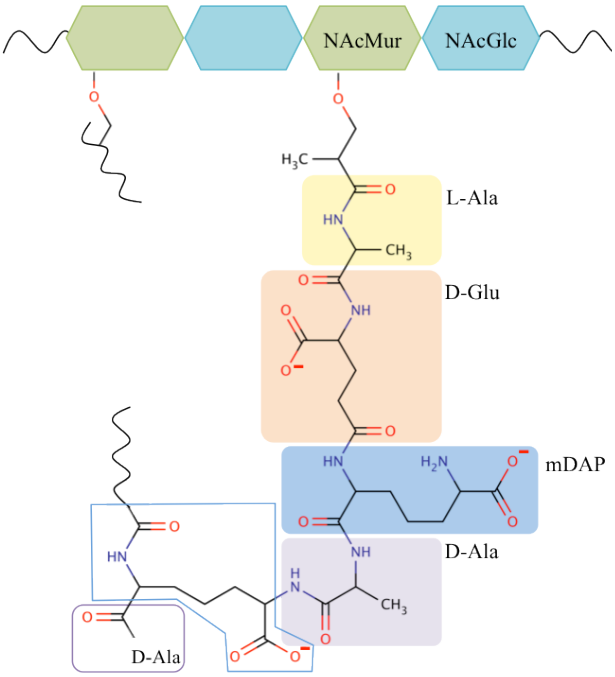


Figure 4.1: Schematic of the PG layer. Two tetrapeptide chains are shown. Peptide components are marked with boxes; filled box for peptide 1 and empty box for peptide 2.

A-E-mDAP-D-Ala—D-Ala. Finally, DacA (D-Ala—D-Ala carboxypeptidase, EC 3.4.17.14) removes the last D-Ala from the peptide chain (Bouhss *et al.,* 2008; Bugg *et al.,* 2011). Figure 4.1 shows a schematic representation of the PG arrangement.

The PG biosynthesis pathway presents a number of potential drug targets and we directed our attention to Ddl. In some bacteria such as *E. coli*, there exist two Ddl proteins (called DdlA and DdlB, Zawadzke *et al.,* 1991) with a sequence identity of about 30% that might result in a lower sensitivity for Ddl inhibitors. There are 587 reviewed Ddl sequences in UniProt (www.uniprot.org), and 18 bacteria presenting two Ddl proteins with sequence identities from 26 to 58%. We are interested in Ddl as new target against *B. pseudomallei* and *F. tularensis* so a search for orthogonal DdlA and DdlB sequences was performed using NCBI BLAST (Johnson *et al.,* 2008). Only one sequence corresponding to DdlB was found in both cases, and as indicated in Part Two, the Dstl labs reported Ddl is essential for *B. pseudomallei*.

The first DdlB structure described was from *E. coli* (*Ec*Ddl, Fan *et al.,* 1997) and another 28, reporting apo and ligand-bound complexes, from various species have been determined. These include the enzymes from *Burkholderia ambifaria* (PDB code 4EG0, Baugh *et al.,* 2013), *B. anthracis* (3R5X), *B. pseudomallei* (4EGQ) and *Y. pestis* (3V4Z). Some of these structures have informed on aspects of mechanism and supported biochemical studies. Two structures of DdlA from *Salmonella typhimurium* have also been determined (3Q1K and QIT2). Ddl presents a fold conserved among different proteins, called the ATP-grasp family in which 21 ligases and kinases are included (Fawaz *et al.,* 2011). See Results and discussion for further details.

The Ddl catalysed reaction follows a sequential ordered mechanism (Prosser and de Carvalho, 2013). ATP binds and the protein conformation changes. The first substrate, N-terminal D-Ala, is phosphorylated and then the C-terminal D-Ala ligated as the phosphate dissociates. Later D-Ala—D-Ala is released followed by ADP (Figure 4.2). The binding order of each D-Ala has not yet been clarified. The mechanism was first proposed by Fan *et al.* (1994) after obtaining the *Ec*Ddl structure in the presence of an intermediate homologue (phosphinate). The reaction mechanism is shared with other ligases from the ATP-grasp family and characterised by the formation of an acylphosphate intermediate. The residues implicated in catalysis are highly conserved between ligases from the ATP-grasp superfamily (Fawaz *et al.,* 2011).

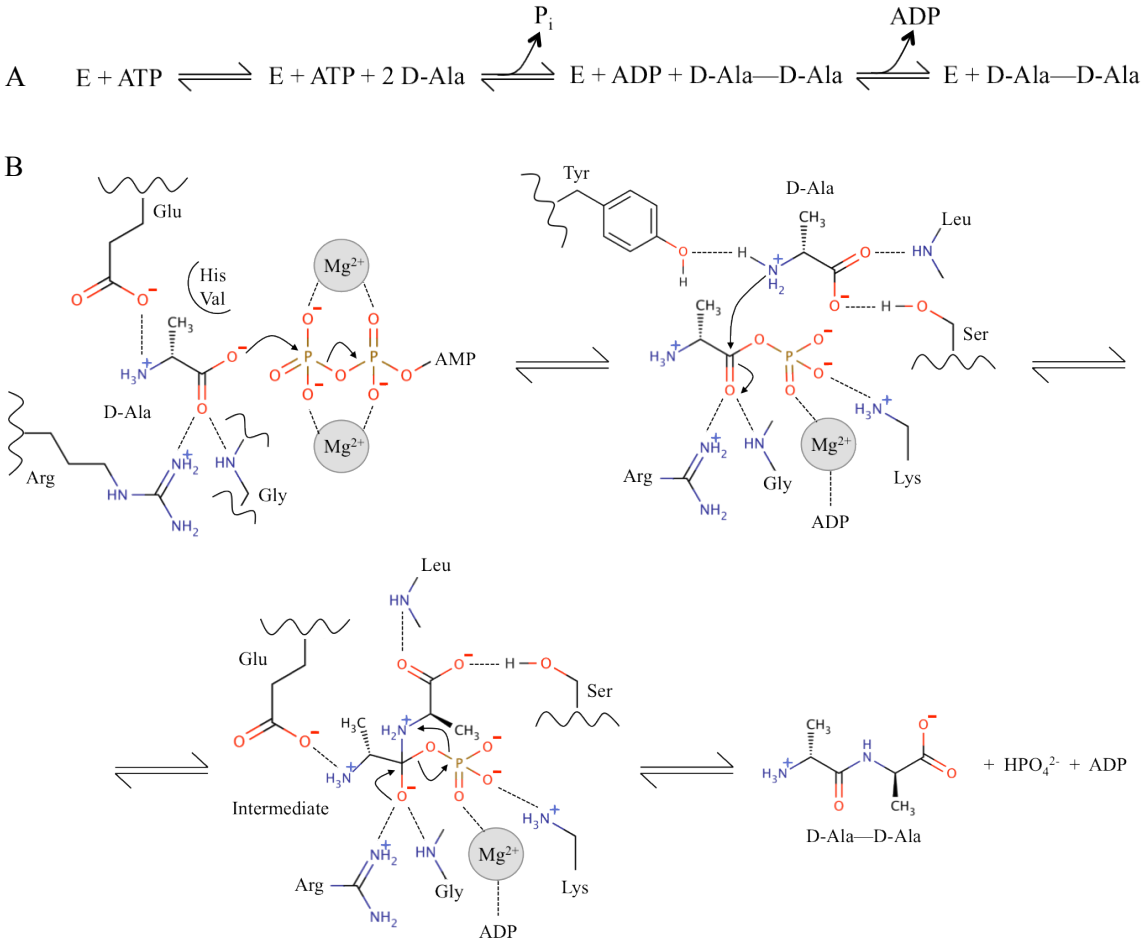


Figure 4.2: A proposed mechanism for Ddl (Fan *et al.,* 1994). (A) Schematic reaction. (B) Mechanism details and amino acids involved.

D-cycloserine (DCS) is a Ddl and AlaR inhibitor (Prosser and de Carvalho, 2013). It is an antibacterial used as a second-line treatment for tuberculosis (Li *et al.,* 2010). However, it is also an agonist of the NMDA (9-*N*-methyl-D-aspartate) receptor (Ollivaux *et al.,* 2014), reduces kynurenic acid synthesis in neurons (Baran and Kepplinger, 2014), and is used to treat some psychiatric disorders (Hofmann *et al.,* 2006). DCS promiscuity, in terms of protein binding, also produces numerous side effects (Li *et al.,* 2010). The DCS K_i values for Ddl range from 2 to 920 μM depending on the targeted organism (BRENDA database, www.brenda-enzymes.org) and MIC (minimum inhibitory concentration) values from 8 to 32 μg ml^{-1} (Putty *et al.,* 2011). There is an *Ec*Ddl structure with a DCS analogue ([(4R)-4-azanyl-4,5-dihydro-1,2-oxazol-3-yl]dihydrogen phosphate, PDB code 4C5A) in the active site pocket. However, the phosphate is located in the ATP γ-phosphate position which suggests the binding might be driven by the anionic component. Therefore, it is not clear how relevant this structure is for the study of DCS binding. The lack of potency of DCS, plus the adverse side effects, indicates this compound might not be an ideal candidate for using as an antibacterial drug targeting Ddl. However it can be a good starting point for the development of new drugs.

Several types of Ddl inhibitors have been developed (Figure 4.3). Parsons *et al.* (1988) developed a series of phosphinates, the first Ddl inhibitors distinct from DCS. The best compounds inhibited *Streptococcus faecalis* Ddl with IC_{50} of 4-50 μM but MIC values for nine different bacteria (4-256 μg ml^{-1}) showed DCS still has more potency (MICs of 4-16 μg ml^{-1}) than phosphinates. Optimisation of these compounds has been carried out (Ellsworth *et al.,* 1996; Fan *et al.,* 1997) but no MIC reported. The most potent phosphinate (in Figure 4.3) binds to the D-Ala pocket and behaves like an irreversible inhibitor. Ddl transfers the γ-phosphate to generate a phosphinophosphate, which has a

dissociation constant of days (Fan *et al.,* 1997). More recently, diazenocarboxamides with Ddl inhibitory activity (IC_{50} 15-50 µM) have been described (Kovač *et al.,* 2007). Determination of MIC values (64-256 µg ml^{-1}) indicates that further improvement is required before these compounds could progress. Sova *et al.* (2009) characterised a hydroxyethylamine as a Ddl inhibitor with an IC_{50} of 110 µM. Recently, D-boroAla and derivatives were described as Ddl inhibitors (Putty *et al.,* 2011). MIC values were reported against various bacteria and show an improvement, compared to DCS, in *Salmonella typhi* (from 64 with DCS to 8 µg ml^{-1}) and *Enterococcus faecium* (from 32 to 16 µg ml^{-1}). They also report a MIC value against *B. pseudomallei* (64 µg ml^{-1}) but no information about DCS inhibition is available. However, the inhibitory constants of D-boroAla compounds have not been performed so Ddl inhibition has not been directly confirmed. Instead of measuring the enzyme activity, the D-Ala and D-Ala—D-Ala concentrations in the presence of various compounds are reported suggesting they are Ddl inhibitors (Putty *et al.,* 2011). An allosteric inhibitor (with K_i values of 20-144 µM for the different protein:substrate complexes) for *Staphylococcus aureus* Ddl has been described and suggests the inhibitory mechanism involves the destabilisation of the

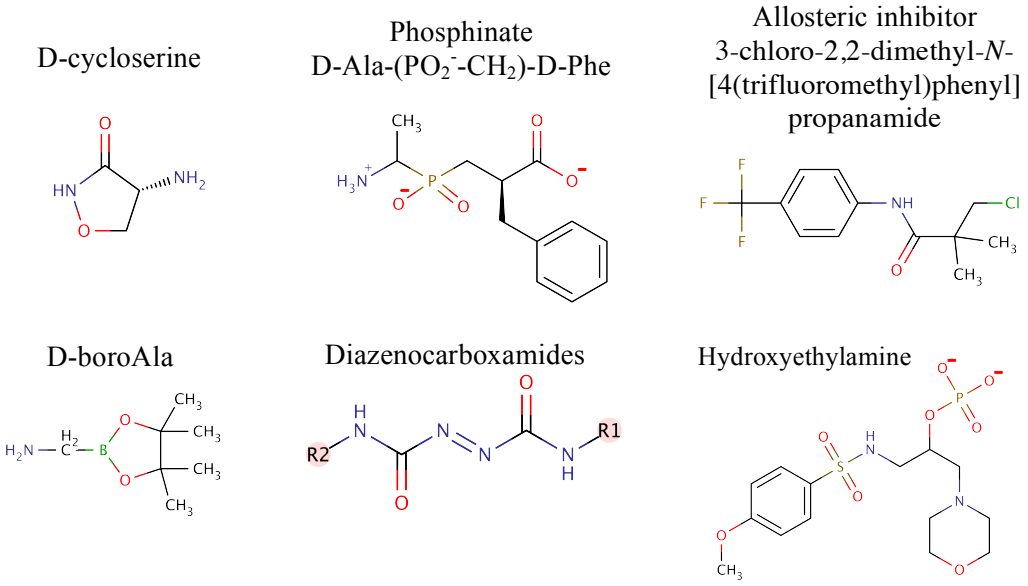


Figure 4.3: Examples of Ddl inhibitors.

reaction intermediate (Liu *et al.,* 2006). Further details on the MIC effect have not been described. A complicating factor in the development of Ddl inhibitors is the fact that these compounds are polar and, consequently, they require active transport for translocation into the cytoplasm.

Despite the early success in finding ligands there has been little progress in developing highly potent inhibitors of Ddl. The lack of clear follow up of the characterised compounds might indicate issues hindering inhibitor validation. The objective here was to assess Ddl as a target for *B. pseudomallei* (*Bp*Ddl) and *F. tularensis* (*Ft*Ddl), and initiate a search for new inhibitor scaffolds. In order to achieve this, it is necessary to produce recombinant protein, obtain crystals in the presence of ligands (including DCS) and establish an assay for validating enzyme inhibition. To search for new inhibitors, an HTP assay was developed and compound library screens conducted.

4.2 Experimental procedures

A flow chart of the experiments carried out is shown in Figure 4.4.

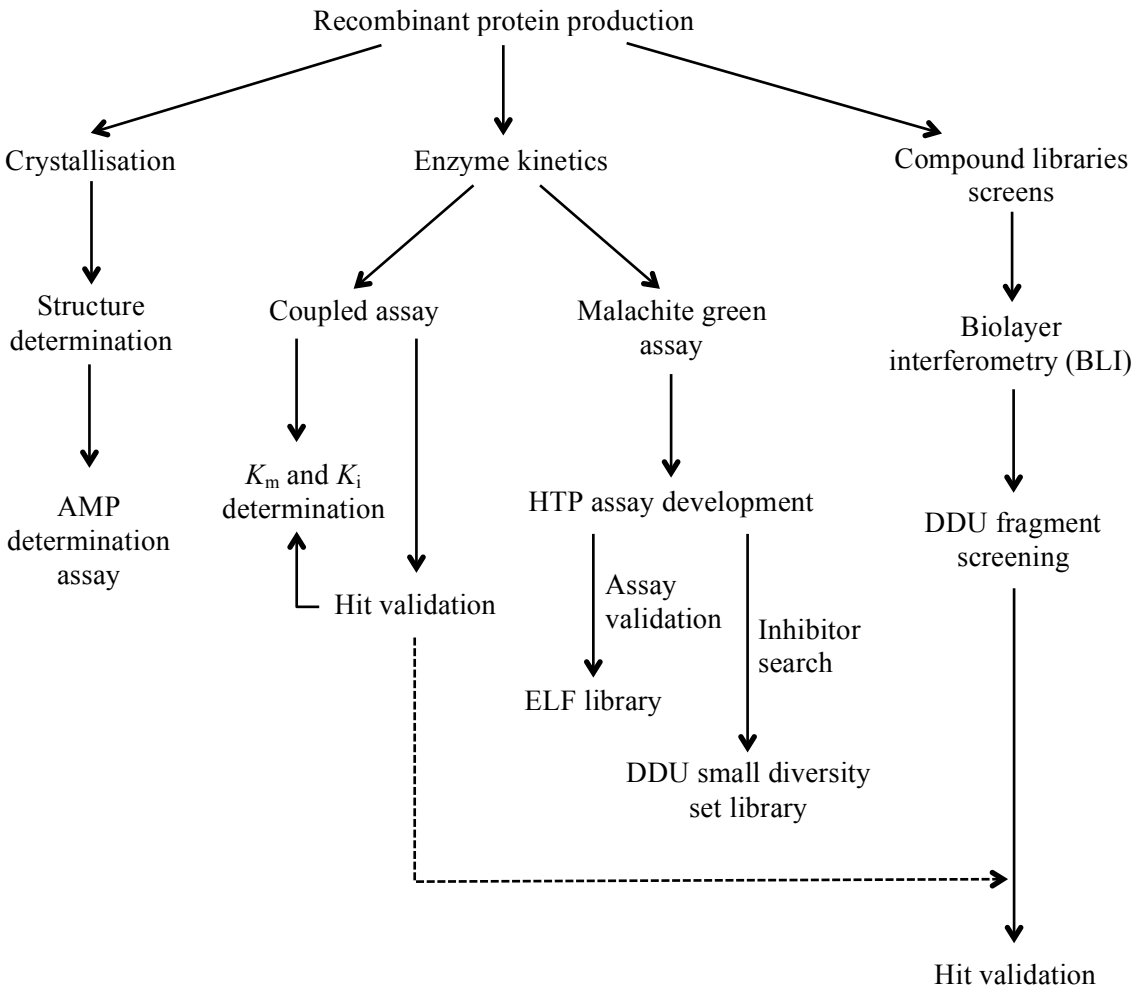


Figure 4.4: Experimental flow chart. Dashed arrow correlates experiments. K_m: Michaelis-Menten constant. K_i: inhibition constant. HTP: high-throughput. ELF: European Lead Factory. DDU: Drug Discovery Unit.

4.2.1 Recombinant protein production

Genes encoding for *Ft*Ddl (UniProt: Q14JP9, protein mass 32 kDa) and *Bp*Ddl (Q63QJ9, protein mass 33 kDa) were purchased from GeneWiz and had 814 bp and 942 bp respectively. The genes were cloned into a pET15bTEV vector using the restriction enzymes NdeI (at the 3' DNA end) and BamHI (5' end) in similar fashion for KynB (section 3.2.1). Cloned vectors were sequenced and gene integrity confirmed.

Recombinant expression was achieved with BL21(DE3) cells, 1 mM IPTG for induction and incubating overnight (O/N) at 20ºC. Purification involved a Ni-NTA column using buffer A (20 mM Tris-HCl pH 7.4, 200 mM NaCl, 5 mM $MgCl_2$) and buffer B (20 mM Tris-HCl pH 7.4, 200 mM NaCl, 5 mM $MgCl_2$, 800 mM imidazole). The His_6-tag was cleaved, during dialysis into buffer A, using 1 mg of TEV per 10 mg of Ddl at 4ºC and incubated O/N. A second Ni-NTA column was performed using the same buffers A and B. To analyse the quaternary structure, SEC, using the HiLoad 16/60 superdex 200 prep grade column (GE Healthcare), was carried out with buffer A and a native-PAGE (Novex). In the case of *Ft*Ddl various quaternary structures were found in SEC. The peak corresponding with a dimer was pooled and prepared for native-PAGE analysis. For crystallisation, two different samples, corresponding to the monomer and dimer peaks, were used. Protein purity and molecular weight were assessed by SDS-PAGE. Protein concentration was determined using the molar extinction coefficients ε(*Ft*Ddl)=29005 M^{-1} cm^{-1} and ε(*Bp*Ddl)=27515 M^{-1} cm^{-1} as calculated in PROTPARAM (Gasteiger *et al.,* 2003).

4.2.2 Crystallographic analysis

4.2.2.1 Protein crystallisation

Protein samples were concentrated to 25 mg ml^{-1}, incubated with 5 mM ATP for 30 minutes on ice, and used for crystallisation trials. Three commercial crystallisation screens were tested using 1:1 and 1:2 protein solution:reservoir ratios. These were The Classics Suite from Qiagen, the JCSG plus screening from Molecular Dimensions, and the PEGs screen also from Molecular Dimensions. Optimisation was performed in hanging drop plates, 1:1 protein solution:reservoir ratio and a final drop volume of 2 µl or 4 µl. Optimisation plates for *Ft*Ddl were set up for conditions H9 JCSG plus and G1 from The Classics Suite, and conditions H12 from The Classics Suite and A2 and H9 from JCSG plus for *Bp*Ddl. Microseeding was used as an additional optimisation step for *Bp*Ddl (Figure 4.5). The seeds were diluted in a solution with the same reservoir components. Three serial 1/5 seed dilutions were performed. The new H9 JCSG plus optimisation plate contained 1:1, 4µl drops and 0.5 µl of the different seed dilutions were added to each drop. For each condition, 4 drops were prepared. The optimisation conditions contained 0.05 to 0.3 M Li_2SO_4, 0.1 M Bis-Tris pH 5.5 and 15-30% (w/v) PEG 3350. Crystals were obtained after co-crystallisation with different ligand combinations using 10 mM for ATP, DCS, D-Ala+ADP, D-Ala—D-Ala+AMP,

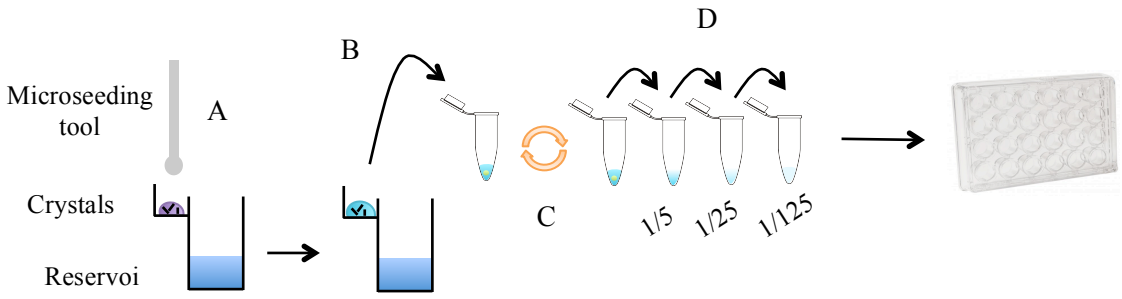


Figure 4.5: Microseeding procedure scheme (A-D). A first step for crystal breakage is done using a glass probe (A, microseeding tool). Then the drop is transferred into an eppendorf tube containing a bead (helps further for fragmentation) and the seeds are vortexed for 30 seconds (B-C). Then, serial 1/5 seed dilutions are prepared (D).

DCS+AMP, DCS+AMPPNP (an non-hydrolysable ATP homologue) and 1 mM of compound #3 which had been identified from a fragment screen. The protein was incubated for 30 minutes in the presence of the different compounds prior to setting up crystallisation.

4.2.2.2 Diffraction and structure determination

Optimised *Bp*Ddl crystals, grown in the presence of ATP, were tested using the in-house X-ray generator (Rigaku M007HF with a Saturn 944HG+ CCD detector). They diffracted to a resolution of 2 to 2.5 Å. A 2.0 Å resolution data set was collected on ESRF beamline BM30-A (*Bp*Ddl:AMP-1). The data were indexed and integrated using XDS (Kabsch, 2010), and scaled and analysed using AIMLESS (Evans, 2011). The structure was solved with PhaserMR (McCoy *et al.,* 2007) using the 2.2 Å resolution apo-*Bp*Ddl structure (PDB code 4EGQ). The model had two subunits in the asymmetric unit. The Matthews coefficient suggested two subunits were present in the *Bp*Ddl:AMP-1 asymmetric unit (V_m of 2.27 $Å^3$ Da^{-1} and bulk solvent of 46%). The apo-*Bp*Ddl dimer was first used for the molecular replacement but no solutions were found. This, together with the different space group (*P*1) that apo-*Bp*Ddl presented, suggested there were structural differences. Finally, the structure was solved using the apo-*Bp*Ddl subunit A. All waters and ligands were removed prior to molecular replacement. A 2.0 Å AMP-bound *Bp*Ddl structure was then obtained. *Bp*Ddl:AMP-1 refinement involved a first round of rigid body refinement followed by cycles of restrained refinement using REFMAC (Murshudov *et al.,* 2011), electron and difference density map checks, and model manipulations made with COOT (Emsley *et al.,* 2010). Other crystals grown in the presence of AMP+D-Ala—D-Ala, AMPPNP+DCS, AMP+DCS, DCS, #3 and without ligands were obtained and tested on the in-house X-ray generator and Diamond Light Source beamline I04-1. A total number of 13 data sets were collected at Diamond.

Data sets were indexed and integrated using XDSgui and iMOLSFLM (Leslie and Powell, 2007), and scaled and merged with AIMLESS. All crystals were isomorphous with *Bp*Ddl:AMP-1. The resolution, R_{merge}, signal-to-noise ratio (<I/σ(I)>), completeness and redundancy ranges for these 13 data sets are summarized in Table 4.1.

	Overall	High resolution shell
High resolution limit (Å)	1.30 - 1.85	
R_{merge}	0.035 - 0.071	0.266 - 0.665
<I/σ(I)>	7.9 - 17.4	1.9 - 4.5
Completeness (%)	94 - 98.6	93.3 - 99.11
Redundancy	2.8 - 4.8	2.8 - 5.0

Table 4.1: Crystallographic statistics summary of data sets from the Diamond Light Source beamline I04-1. $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and <I(*hkl*)> is the mean value of $I_i(hkl)$ for all *i* measurements.

The structure was determined for a crystal grown in the presence of ADP and D-Ala (1.3 Å resolution). Molecular replacement was performed with the refined subunit A from *Bp*Ddl:AMP-1. The results showed the presence of AMP and this structure was named *Bp*Ddl:AMP-2. Molecular replacement using the refined *Bp*Ddl:AMP-2 dimer was carried out to solve the structure at 1.5 Å from a crystal containing AMP and D-Ala—D-Ala. In this case, both AMP and D-Ala—D-Ala were present and named *Bp*Ddl:D-Ala—D-Ala. Refinement for both structures involved tight non-crystallographic symmetry (NCS) restraints at the beginning, which were released after the first refinement cycles. For the three structures, side-chain rotamers, water molecules and ligands were added when the protein models were completed. Given that high resolution diffraction data were available, anisotropic thermal factors were included in the refinement. Refinement was finished when no improvements were observed in model geometry and the R_{free} and R_{work} values, and density maps indicated no further changes were justified. For the other 11 structures, omit maps were inspected after one cycle of refinement using the *Bp*Ddl:AMP-2 dimer model. AMP was present in all samples, even when no ligands were added. This indicated AMP might be present

in the protein solution. To investigate this, determination of the AMP in the samples was carried out (see section 4.3.4.2). There was no reliable electron density for some residues in the models and they were not included. The absent amino acids are: 1) residues 1-3, 153-155 (subunit A) and 1-8, 153-155 (subunit B) in *Bp*Ddl:AMP-1 model, 2) residues 1-2 (subunit A) and 1-3 (subunit B) in *Bp*Ddl:AMP-2, and 3) residues 1 (subunit A) and 1-3 (subunit B) in *Bp*Ddl:D-Ala—D-Ala.

4.2.2.3 Ligand identification

All crystals were harvested form drops containing Li_2SO_4, $MgCl_2$, NaCl, Tris-HCl, Bis-Tris, PEG 3350 and the ligand combinations described above. This was taken into account when making decisions about the presence of ligands in the models.

Difference Fourier maps at the AMP site clearly indicated this compound was present in all structures. Substitutions of AMP by ADP were tried in *Bp*Ddl:AMP-1 and *Bp*Ddl:AMP-2. This showed strong negative electron density (about 14 σ level) at the β-phosphate of ADP inferring AMP is the molecule present. In the case of *Bp*Ddl:D-Ala—D-Ala, a difference Fourier map suggested the presence of D-Ala—D-Ala. If full occupancy was assigned for all the ligand atoms, then negative electron density was noted at the N-terminus. However, if only one D-Ala was placed, a positive feature was observed indicating a bigger molecule was present. According to this, D-Ala—D-Ala was added and zero occupancy was assigned for the N-terminal amino and methyl groups.

Additionally, the difference Fourier maps suggested the presence of other entities in both *Bp*Ddl:AMP-2 and *Bp*Ddl:D-Ala—D-Ala. To investigate this, eight sulfur atoms in the asymmetric unit were removed to calculate their positive density from the difference

Fourier map. The average *B*-factors of these sulfurs were 17.3 $Å^2$ in *Bp*Ddl:AMP-2, and 16.7 $Å^2$ in *Bp*Ddl:D-Ala—D-Ala. From the difference Fourier map, average values of 1.9 σ per electron (*Bp*Ddl:AMP-2) and 1.8 σ per electron (*Bp*Ddl:D-Ala—D-Ala) were calculated. These values were used to estimate the number of electrons at the sites of interest.

The ion sites close to His70 (subunit A) presented values of 28.1 σ (14 electrons, *Bp*Ddl:AMP-2) and 22.7 σ (12.6 electrons, *Bp*Ddl:D-Ala—D-Ala). When Mg^{2+} ions were added, they refined with *B*-factors of 18 $Å^2$ and 22 $Å^2$. The average coordinate bond length with waters is 2.09 Å in *Bp*Ddl:AMP-2 and 2.11 Å in *Bp*Ddl:D-Ala—D-Ala. The coordination bond distances to the N from His70 are 2.14 Å and 2.10 Å respectively. These values are close to the expected coordinate bond lengths of 2.07 Å for waters and 2.05 Å for N from histidines (Harding, 2001). They presented a coordination geometry of 6 that is characteristic of Mg^{2+} ions. All this confirmed the presence of Mg^{2+}.

Two other difference peaks were observed in *Bp*Ddl:AMP-2 (near residues 293 from subunit A, 221 from subunit B) and three in *Bp*Ddl:D-Ala—D-Ala (near residues 293 from subunit A, 221 and 290 from subunit B). Firstly, the positive density sites in *Bp*Ddl:D-Ala—D-Ala were characterised. They presented values of 18.9, 11.3 and 16.5 σ (11, 6 and 9 electrons respectively), and the same tetrahedral-like shape. These sites did not show evidence (*e.g.* additional positive density peaks corresponding to coordinating-water molecules) of a coordination number of 6 so Mg^{2+} was discarded as the probable ligand. When Cl^- was placed, positive electron density around the Cl^- and average *B*-factor value of 45 Å were observed. Replacement of Cl^- by SO_4^{2-} molecules made the *B*-factor decrease (average of 26.7 Å) and no positive density was present.

Negative electron density was observed at one of the sites and an occupancy of 0.5 was assigned. The positions of the two difference peaks in *Bp*Ddl:AMP-2 were conserved with the *Bp*Ddl:D-Ala—D-Ala structure and SO_4^{2-} ions were also placed with 0.5 occupancy (*B*-factors average of 30 Å). The SO_4^{2-} molecules interact with N (from, Arg, Asn, Gly and Lys), and O (Tyr, water) atoms, that can behave as hydrogen bond donors. The SO_4^{2-} ions located near residues 221 are in the D-Ala binding pocket. Comparing the models with the phosphinate-bound *E. coli* Ddl structure (PDB code 1IOV), the SO_4^{2-} ions are located in a different position than the phosphate of the possible reaction intermediate.

4.2.2.4 Further structural analysis

To assess the geometry of the models MolProbity (Chen *et al.,* 2010) was used. Secondary-structure was determined with DSSP (Touw *et al.,* 2015), dimer formation analysed using PISA (Krissinel and Henrick, 2007), and *B*-factors analysed using Baverage from the CCP4i suite. Figures were prepared using ALINE (Bond and Schüttelkopf, 2009), PyMOL (www.pymol.org) and Molsoft ICM (www.molsoft.com). The DALI server (Holm and Rosenstrom, 2010) was used to search for structural homologues. Electrostatic surfaces were determined using Molsoft ICM.

A comparison of the protein sequences of all Ddl orthologues was carried out. First, a search for relevant sequences in UniProt (www.uniprot.org) was conducted. The identified sequences were grouped by 50% identity and aligned using Clustal Omega (Sievers *et al.,* 2011) with group representatives (588 sequences).

Analysis of the protein dimer interface was carried out using Molsoft ICM. First, a C_α overlay of the available Ddl structures with *Bp*Ddl:AMP-2 was performed. Then, using

the "contact areas" tool, residues at the dimer interface were identified. This was used to analyse the conservation of these amino acids in different organisms.

The druggability score, described previously, predicts the likelihood of a pocket to bind drug-like compounds. The pocket finder tool from the Molsoft ICM defines pockets from a three-dimensional structure and provides a druggability score for each. This was used to calculate the theoretical druggability of *Bp*Ddl pockets in the presence and absence of the co-factor. When the score is greater than 0.5 the pocket is considered druggable. Two proteins were used as controls, dihydrofolate reductase (DHFR, Keedy *et al.,* 2014, PDB code 4PSS) and neuraminidase (Vavricka *et al.,* 2013, PDB code 4H53). DHFR presents a well-defined druggable pocket while neuraminidase has highly polar pockets difficult to target and is considered as non-druggable-like by Molsoft ICM.

4.2.3 AMP detection

The crystallographic analysis revealed that AMP was bound to the protein even when no nucleotide was added. Therefore, the AMP-Glo kit from Promega was used to quantify the presence of AMP in the purified *Bp*Ddl sample. This assay measures the presence of AMP using a system of coupled enzymes. The first one, adenylate kinase (AK), transforms AMP to ADP by transferring the γ-phosphate from dCTP. A second enzyme, nucleoside-diphosphate kinase (NDK), transfers the γ-phosphate from dCTP to ADP and forms ATP. The enzyme luciferase oxidises the substrate luciferin hydrolysing the ATP releasing AMP, pyrophosphate, CO_2 and light (Figure 4.6).

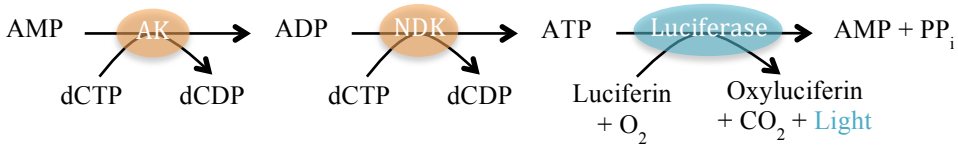


Figure 4.6: AMP-Glo kit reactions.

The reaction was performed in a 96-well white plate, using 25 µl of sample, in triplicate and with the layout described in Figure 4.7. The addition of the different reagents and the standard curve preparation were carried out as indicated in the kit manual. The signal was measured with the FLUOstar OPTIMA microplate reader from BMG Labtech.

The control protein, selected because it does not bind AMP/ADP/ATP, was *Ba*KynB. Samples were pre-treated for denaturation before performing the AMP determination assay. Samples treatments included variations in the temperature and pH, and digesting the proteins with chymotrypsin. On one hand, *Bp*Ddl and *Ba*KynB samples were incubated for 20 minutes at 100ºC prior to the assay performance. In parallel, the pH of the samples was first decreased to 2 by the addition of HCl. After incubating the samples for 2 minutes, NaOH was added and a pH of 10 was obtained. Again, incubation was for 2 minutes and HCl used to reach a pH value of 7. The samples used for chymotrypsin digestion were *Bp*Ddl, *Ba*KynB and *Ba*KynB in the presence of 2 µM AMP (*Ba*KynB-AMP). First the samples were incubated for 20 minutes at 100ºC.

	1	2	3	4	5	6	7	8	9	10	11	12
A	10	5	2.5	1.25	0.63	0.31	0.16	0.08	0.04	0.02	0.01	0
B	WO	KynB WT	Ddl WT	KynB ΔpH	Ddl ΔpH	KynB ΔT	Ddl ΔT	KynB Ch	KynB Ch+	Ddl Ch		
C	10	5	2.5	1.25	0.63	0.31	0.16	0.08	0.04	0.02	0.01	0
D	WO	KynB WT	Ddl WT	KynB ΔpH	Ddl ΔpH	KynB ΔT	Ddl ΔT	KynB Ch	KynB Ch+	Ddl Ch		
E	10	5	2.5	1.25	0.63	0.31	0.16	0.08	0.04	0.02	0.01	0
F	WO	KynB WT	Ddl WT	KynB ΔpH	Ddl ΔpH	KynB ΔT	Ddl ΔT	KynB Ch	KynB Ch+	Ddl Ch		
G												
H												

Figure 4.7: Plate layout for the AMP detection experiment. **WO** wells contain only buffer (without protein or AMP). In **blue** are the standard curve samples and the numbers indicate the AMP contractions (µM). **KynB** is the sample *Ba*KynB, and **Ddl** is *Bp*Ddl. **WT** indicates the samples where not treated. **ΔT**, **ΔpH**, **Ch** and **Ch+** are used when samples were under temperature and pH stress or digested with chymotrypsin. + indicates 2 µM AMP was added to the sample.

Chymotrypsin was added to a final ratio chymotrypsin:sample of 1:20 and the mixture was incubated O/N at 25ºC. Then, the samples were filtered using an Amicon ULTRA 3K concentrator. The protein concentration in each well was 4 mg ml^{-1} for *Ba*KynB and 25 mg ml^{-1} for *Bp*Ddl.

4.2.4 Enzyme kinetics: coupled enzymatic assay

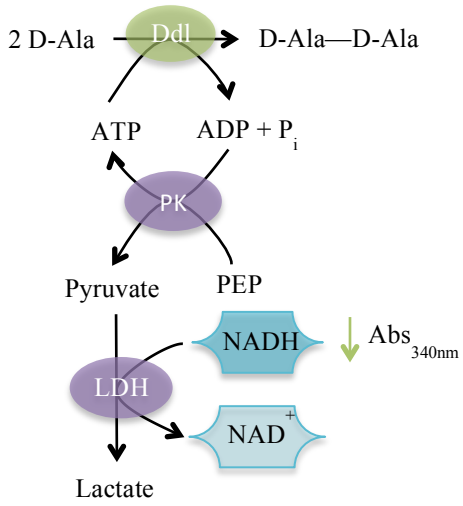


Figure 4.8: Enzymatic assay scheme. Ddl (*B. pseudomallei* Ddl), PK (pyruvate kinase), LDH (lactate dehydrogenase), PEP (phospho-enol pyruvate), Abs (absorbance).

The activity of *Bp*Ddl was measured using a spectrophotometric assay (Prosser and de Carvalho, 2013). The formation of ADP was measured by a pyruvate kinase/ lactate dehydrogenase (PK/LDH) coupling system (Figure 4.8). PK carries out the first coupled reaction, a transfer of the phosphate from the phospho-enol-pyruvate (PEP) to the ADP produced by Ddl to form pyruvate and ATP. Next, LDH reduces pyruvate to lactate by NADH oxidation. The decrease in the absorbance at 340 nm due to NADH oxidation is recorded. Measurements were made using a UV-2450 Shimadzu spectrophotometer. The decrease in NADH concentration can be quantified using the ε_{340nm}(NADH) = 6220 $M^{-1}cm^{-1}$ and the Lambert-Beer law.

The Ddl catalysed reaction presents a sequential ordered mechanism and the equations used are described in Prosser and de Carvalho, 2013 (Equation 4.1).

$$V = k_{cat} \frac{[ATP]\,[D\text{-}Ala]}{K_{i,ATP}\,K_{m2} + K_{m2}[ATP] + K_{m,ATP}[D\text{-}Ala] + [ATP]\,[D\text{-}Ala]}$$

Equation 4.1: General equation for Ddl mechanism (Prosser and de Carvalho, 2013). $K_{i,ATP}$ and $K_{m,ATP}$ are the inhibition and Michaelis-Menten constants for ATP, K_{m2} the Michaelis-Menten constant for the second D-Ala (considering K_{m1} from the first D-Ala much smaller than K_{m2}), and k_{cat} the catalytic constant.

4.2.4.1 D-alanine K_m determination

The assay buffer contained 100 mM Tris-HCl pH 8, 100 mM KCl, 10 mM $MgCl_2$, 2.5 mM PEP, 0.4 mM NADH, 5 µl ml^{-1} PK/LDH mixture from SIGMA, 30 µM *Bp*Ddl, 500 µM ATP and variable concentrations of D-Ala (0.1, 0.3, 0.5, 0.7, 1, 1.5, 2, 2.5, 3, 5, 10, 20 and 100 mM). Ten times (10x) concentrated D-Ala stocks were prepared for every D-Ala assay concentration so the volume added was always kept constant (10 µl). The final reaction volume was 100 µl. A master mix was prepared containing all assay buffer components except the D-Ala and the enzyme. All components were incubated for 30 minutes at 30ºC and then the assay performed in triplicate. Negative controls included master mix with *Bp*Ddl and without substrate, and master mix without *Bp*Ddl and with D-Ala. No decrease in the absorbance was detected in the negative controls during the assay period. The positive control was the PK/LDH system and instead of ATP, 0.4 mM ADP was added.

(A) $$V = \frac{k_{cat}\,[S]}{K_{m2} + [S]}$$ (B) $$\frac{1}{V} = \frac{1}{k_{cat}}\left(1 + \frac{K_{m2}}{[S]} + \frac{K_{m1}\,K_{m2}}{[S]^2}\right)$$

Equation 4.2: (A) Michaelis-Menten equation. (B) Double reciprocal equation for K_{m1} calculation. V = reaction velocity (nmol/min), K_{m1} = Michaelis-Menten constant for the first D-Ala (M), K_{m2} = Michaelis-Menten constant for the second D-Ala (M), V_{max} = maximum velocity of the reaction (nmol/min), [S] = variable substrate concentration (M).

Initial velocities were obtained by measuring the steady-state rate for 60 seconds. The absorbance data were plotted against time (seconds) and adjusted using a linear

regression analysis. The initial velocities were obtained from the slope of the regression lines and plotted against the concentration of D-Ala. At a constant and saturating concentration of ATP, there exist two K_m; one for the first D-Ala to bind (K_{m1}) and another for the second (K_{m2}). For their determination different equations were used. Firstly, the k_{cat} and K_{m2} values were obtained under the assumption that K_{m1} is much smaller than K_{m2}. According to this, the general velocity equation will be the Michaelis-Menten equation (Equation 4.2A, Jonson and Goody, 2011). The maximum velocity of the reaction (V_{max}) can then be calculated from:

$$k_{cat} = V_{max} / [E]_T, \text{ where } [E]_T \text{ is the total enzyme concentration.}$$

The K_{m2} and k_{cat} values can be included in the reciprocal plot Equation 4.2B to calculate the K_{m1} at low D-Ala concentration (0.1 to 1 mM).

Determination of K_{m1} and K_{m2} values was also carried out at pH 7.5 using D-Ala concentrations of 0.1, 0.5, 1 1.5, 2, 2.5, 3, 5 and 20 mM. Data analysis was performed as described above.

4.2.4.2 ATP K_m determination

The assay buffer contained 100 mM Tris-HCl pH 8, 10 mM KCl, 10 mM $MgCl_2$, 2.5 mM PEP, 0.4 mM NADH, 5 µl ml^{-1} PK/LDH mix from SIGMA, 10 mM D-Ala, 20 µM *Bp*Ddl and various ATP concentrations (0.1, 0.2, 0.5, 1.2 and 3 mM). ATP stocks were prepared at 10x the final assay concentration. A master mix was prepared containing all assay buffer components except for ATP and enzyme. The same controls and experimental procedures as in K_m(D-Ala) determination experiments were carried out. Initial velocities were analysed using the substrate inhibition equation (Equation 4.3), included in GraphPad Prism (www.graphpad.com).

$$V = \frac{V_{max}\,[S]}{K_{m,ATP} + [S]\,(\,1 + [S]/K_i\,)}$$

Equation 4.3: Substrate inhibition. V = reaction velocity (nmol/min), $K_{m,ATP}$ = Michaelis-Menten constant (M) for ATP, V_{max} = maximum velocity of the reaction (nmol/min), [S] = ATP concentration (M).

4.2.4.3 D-cycloserine and fosmidomycin inhibition assays

Experiments were performed in the presence of different inhibitor concentrations to determine K_m(D-Ala). In the case of DCS the concentrations used were 100, 200, 250, 300, 500 and 1000 μM. For fosmidomycin, concentrations of 1, 10, 100, 500, 1000 μM were tested. Stocks (10x) for every inhibitor final concentration were prepared (see section 4.2.8.2 for compound structures and source-stock concentration). Each K_i determination experiment was conducted for about 12 hours. To check for loss of *Bp*Ddl activity during the assay, an experiment without inhibitors was performed before and after the inhibition assay.

K_i calculations were performed using double reciprocal plots (Equation 4.4A). This will inform about the mechanism of inhibition. The linearised data were fitted into linear regression lines. The slopes of these lines were then plotted against the concentration of DCS (Equation 4.4B) and data were again fitted into a linear regression curve. The intercept with the y-axis gives the value of K_m/V_{max} and the slope the $K_m/(V_{max}Ki)$.

(A) $$\frac{1}{V} = \frac{1}{V_{max}} + \frac{K_m}{V_{max}}\left(1 + \frac{[I]}{K_i}\right)\left(\frac{1}{[S]}\right)$$

(B) $$\text{Slope} = \frac{K_m}{V_{max}} + \frac{K_m}{V_{max}K_i}\,[I]$$

Equation 4.4: (A) Double reciprocal equation for competitive inhibition. (B) Secondary representation for inhibitory data. V = reaction velocity (nmol/min), K_m = Michaelis-Menten constant (M), V_{max} = maximum velocity of the reaction (nmol/min), [S] = substrate concentration (M), [I] = inhibitor concentration, K_i = inhibitory constant.

Rearrangement of these equations allows calculation of the K_i (Prosser and de Carvalho, 2013).

To check for inhibition of any of the coupled enzymes, a cross-validation assay was performed. The assay buffer in this case contained 100 mM Tris-HCl pH 8, 100 mM KCl, 10 mM $MgCl_2$, 2.5 mM PEP, 0.4 mM NADH, 5 µl ml^{-1} PK/LDH mixture from SIGMA and 0.5 mM ADP. Concentrations of 0.5 and 10 mM DCS (from stock 1M) and a positive control (without inhibitor) were tested. Calculation of velocity rates were performed as in section 4.2.4.1.

4.2.4.4 Ddl activity at different pH values

*Bp*Ddl activity at different pH values was tested. In this case, the reaction mixture contained 100 mM of the buffer (Table 4.2), 10 mM KCl, 10 mM $MgCl_2$, 2.5 mM PEP, 0.4 mM NADH, 5 µl ml^{-1} PK/LDH mix from SIGMA, 0.5 mM ATP, 10 mM D-Ala, and 20 µM *Bp*Ddl (added for starting the reaction). The buffers used for the different pH values are annotated in Table 4.2. The maximum reaction rate was assigned as 100% activity.

Buffer	pH	Buffer	pH
Bis-Tris	5.5	Tris-HCl	8.0
Tricine	5.1	Tris-HCl	8.5
MES	6.0	HEPES	8.5
HEPES	7.5	BICINE	9.0
Tris-HCl	7.5	CHES	9.5

Table 4.2: Buffers used for the determination of the pH effect curve.

4.2.5 Enzyme kinetics: malachite green assay and high-throughput assay development

Another detection method used for characterising *Bp*Ddl was the malachite green assay. The reagents were purchased from Enzo Life Sciences as a ready-to-use mixture, called BIOMOL Green. The mixture contains molybdate (MoO_4^{2-}) and malachite green (organic dye). MoO_4^{2-} binds to the released phosphate generating the MoO_4^{2-}:P_i complex. Then malachite (yellow) binds to MoO_4^{2-}:P_i and produces a green malachite:MoO_4^{2-}:P_i complex that absorbs at 620 nm (Pegan *et al.,* 2010). The acidic nature of this mixture (pH about 1) necessitates an end-point assay.

A HTP assay was developed for screening compound libraries (sections from 4.2.5.1 to 4.2.5.5). The BIOMOL Green assay was used to determine the optimum enzyme concentration, K_m and a time course experiment to check for linearity. Two compound libraries were screened. The first one was obtained from the European Lead Factory (ELF, www.europeanleadfactory.eu), a partnership between public and private organisations. The HTP assay development and the ELF library screen, which contains a selection of compounds from three different commercially available sources, were conducted at the ELF in NewHouse, Scotland. These are the NCC (NIH Clinical Collection, www.nihclinicalcollection.com) library that contains known drugs in clinical trials, the Selleck Bioactive screen (www.selleckchem.com) containing known drugs, natural products and known protein inhibitors, and the BioAscent library comprising a lead-like small molecule collection. The second library screen was with the DDU (Drug Discovery Unit, www.drugdiscovery.dundee.ac.uk) small diversity set. It consists of a collection of compounds that covers a wide range of the chemical space. The plates for the HTP assay were clear bottom black/white 384 well plates from Greiner (code 781091 for black and 781101 for clear plates). For the ELF screen, black

plates were used and clear plates for the diversity set screening. The reactions were performed at room temperature and in a final volume of 30 µl.

4.2.5.1 Optimisation of enzyme concentration

The assay buffer contained 100 mM Tris-HCl pH 8, 10 mM KCl and $MgCl_2$ and final substrate and co-factor concentrations of 10 mM D-Ala and 0.25 mM ATP. The assay solutions were prepared as follows:

- Assay buffer (B)
- Protein solution 2x (P2x): 60 nM *Bp*Ddl in B
- Substrate solution 2x (S2x): 20 mM D-alanine, 0.5 mM ATP in B

The plate was prepared adding 18 µl P2x in row A and assay buffer in the rest of the rows. Serial dilutions were performed (3 µl from $row_{A/n}$ to row_{n+1}) from row A to P in a final volume of 15 µl. Row P was the "without enzyme" control.

To start the reaction, 15 µl of S2x were added to every well at different time points per column (0, 5, 10, 15, 17, 19, 19.5 and 20 min) in triplicate as indicated in Table 4.3. For the last 3 columns, the reaction was stopped before adding the S2x.

Column	t_{add} (min)	$t_{incubation}$ (min)	Column	t_{add} (min)	$t_{incubation}$ (min)
1	0	20	13	17	3
2	0	20	14	17	3
3	0	20	15	17	3
4	5	15	16	19	1
5	5	15	17	19	1
6	5	15	18	19	1
7	10	10	19	19.5	1.5
8	10	10	20	19.5	1.5
9	10	10	21	19.5	1.5
10	15	5	22	20	0
11	15	5	23	20	0
12	15	5	24	20	0

Table 4.3: Assay plate time points. t_{add} = time when the reaction was started. $t_{incubation}$ = length of incubation before stopping the reaction.

Next, 30 µl of BIOMOL Green were added to each well simultaneously. Due to the low pH of the final mixture, non-enzymatic hydrolysis of ATP can occur and the released phosphate interacts with MoO_4^{2-} to increase the background signal. To avoid this, after incubating the mixture for 5 minutes, 5 µl of 20% (w/v) tri-sodium citrate were added to every well. Citrate forms a complex with free MoO_4^{2-}:P_i preventing further binding of malachite (Aherne *et al.,* 2003). Prior to the absorbance measurement, the reaction was incubated for 20 minutes. The final plate layout is given in Figure 4.9.

		1	2	3	4	5	6	7	8	9	21	22	23	24
	[ENZ] nM // min	20	20	20	15	15	15	10	10	10	0.5	0	0	0
A	30													
B	5													
C	8.3E-01													
D	1.4E-01													
O	3.8E-10													
P	0.0E+00													

Figure 4.9: Layout of the plate for the enzyme titration experiment. Every row contains a different protein concentration and every column was used for a specific time point.

Absorbance was recorded at 620 nm using on EnVision plate reader from PerkinElmer. Data were analysed using Microsoft EXCEL. Absorbance was plotted against time for each protein concentration to calculate the initial velocities.

4.2.5.2 D-alanine K_m determination

The different buffers prepared for the D-Ala K_m determination are:

- Buffer B: 100 mM Tris-HCl pH 8, 10 mM KCl, 10 mM $MgCl_2$
- Protein solution 2x (P2x): 0.4 nM *Bp*Ddl in B
- Substrate solution 2x (S2x): 20 mM D-alanine and 1 mM ATP in B
- Assay buffer (B2x): B and 1 mM ATP

Variations in the D-Ala concentration were set up by rows, and different time points in columns (10 sec, 30 sec, 1, 1.5, 2, 3, 5, 7, 10, 15, and 20 min). First 20 µl of S2x were

added in rows A and B, and 15µl of B2x from rows C to O. Serial dilutions were made by adding 5 µl from row$_{A/odd\ number}$ to row$_{odd\ number+1}$, and from row$_{B/even\ number}$ to row$_{even\ number+1}$. Columns 1 and 2 were filled with 15 µl of B instead of P2x (negative control, without enzyme). Row P did not contain the substrate D-Ala. The layout of the plate is shown in Figure 4.10.

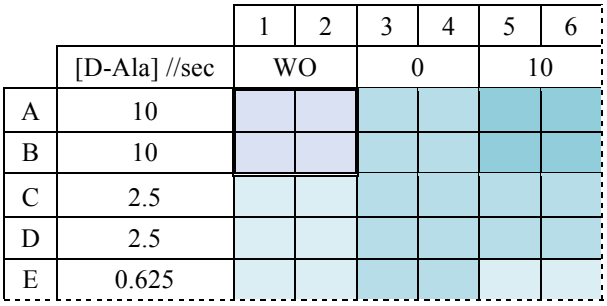


Figure 4.10: Plate layout for the K_m determination experiments. There are four replicates for every time point (the double line square shows the replicates for the first time point). **WO** indicates *Bp*Ddl was not present in the well.

The reaction was started by adding 15 µl of P2x to columns 3 to 24 at different time points (Table 4.4).

The addition of BIOMOL Green and tri-sodium citrate was performed as in section 4.2.5.1.

Column	t_{add} (min)	$t_{reaction}$ (sec)	Column	t_{add} (min)	$t_{reaction}$ (sec)
1	WO	0	13	17	180
2	WO	0	14	17	180
3	20	0	15	15	300
4	20	0	16	15	300
5	19.2	10	17	13	420
6	19.2	10	18	13	420
7	19.5	30	19	10	600
8	19.5	30	20	10	600
9	19	60	21	5	900
10	19	60	22	5	900
11	18	120	23	0	1200
12	18	120	24	0	1200

Table 4.4: Assay plate time points. t_{add} = time when the reaction was started. $t_{incubation}$ = how long the reaction was incubated before stopping it. **WO** indicates *Bp*Ddl was not present in the well.

Data analysis was carried out using Microsoft EXCEL for initial velocity calculations. In this case, only calculation of K_{m2} was performed. GraphPad Prism was used for K_{m2} determination using the Michaelis-Menten equation.

K_{m2} in the presence of a lower ATP concentration was calculated following the protocol described in this section but using a fixed ATP concentration of 0.2 mM instead of 0.5 mM.

4.2.5.3 ATP K_m determination

A similar buffer preparation was set up as used for D-Ala K_m determination:

- Buffer B: 100 mM Tris-HCl pH 8, 10 mM KCl, 10 mM $MgCl_2$
- Protein solution 2x (P2x): 0.4 nM *Bp*Ddl in B
- Substrate solution 2x (S2x): 1 mM ATP and 20 mM D-Ala in B
- Assay buffer (B2x): B and 20 mM D-Ala

Variations in ATP concentrations were set up by rows and different time points in columns (10 sec, 30 sec, 1, 1.5, 2, 3, 5, 7, 10, 15, and 20 min). First 20 µl of S2x were added in rows A and B, and 15 µl of B2x from rows C to P. Serial 1/4 dilutions were made by adding 5 µl from $row_{A/odd\ number}$ to $row_{odd\ number+1}$, and from $row_{B/even\ number}$ to $row_{even\ number+1}$. Columns 1 and 2 were filled with 15 µl of B instead of P2x (negative control, without enzyme). Row P did not contain the substrate D-Ala. The layout of part of the plate is represented in Figure 4.11.

The reaction was started by adding 15 µl of P2x to columns 3 to 24 at different time points (Table 4.5, section 4.2.5.2).

		1	2	3	4	5	6
	[ATP] //sec	WO		0		10	
A	0.75						
B	0.75						
C	0.187						
D	0.187						
E	0.047						

Figure 4.11: Plate layout for the K_m determination experiments. There are four replicates for every time point (the double line square shows the replicates for the first time point). **WO** indicates *Bp*Ddl was not present in the well.

The addition of BIOMOL Green and tri-sodium citrate was performed as explained in section 4.2.5.1, and data analysis as in section 4.2.5.2.

4.2.5.4 Time course experiment

To determine whether the assay was linear under the chosen assay conditions, a time course experiment was carried out. Assay solutions included:

- Buffer B: 100 mM Tris-HCl pH 8, 10 mM KCl, 10 mM $MgCl_2$
- Protein solution 2x (P2x): 0.4 nM *Bp*Ddl in B
- Substrate solution 2x (S2x): 400 µM ATP and 10 mM D-Ala in B

The reaction was started at time points 2.5 minutes apart during 90 minutes (see plate layout in Figure 4.12). In this case, there are 8 replicates for each time point.

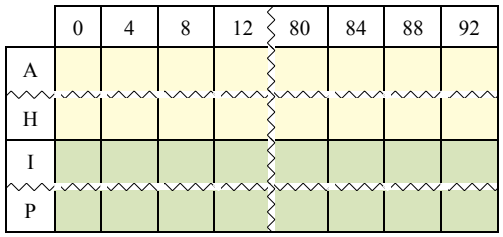


Figure 4.12: Time course experiment plate layout. Numbers indicate the reaction time (min) for each column.

Calculations of the Z' score (Equation 4.5) and the signal-to-noise ratio were performed using Microsoft EXCEL.

$$Z' = 1 - \frac{3\,(s.d._{max} + s.d._{min})}{mean(max) - mean(min)}$$

Equation 4.5: Z' score equation. s.d = standard deviation, max = maximum signal, min = minimum signal.

4.2.5.5 Determination of DMSO tolerance and D-cycloserine IC_{50}

The assay solutions used in this experiment were:

- Buffer B: 100 mM Tris-HCl pH 8, 10 mM KCl, 10 mM $MgCl_2$
- Protein solution 3x (P3x): 0.6 nM *Bp*Ddl in B
- Substrate solution 3x (S3x): 600 µM ATP and 15 mM D-Ala in B
- Compounds (CPD) solutions 3x:

- DCS solution A (DCS3x_A): 1 mM DCS in B
- DCS solution B (DCS3x_B): 0.1 mM DCS in B
- DMSO solution (DMSO3x): 80% DMSO in B

Three different plates were set up. Plate 1 (CPD plate, Figure 4.13) was divided into four different sections. The first addressed the assay in the absence of any inhibitor. The second and third were prepared using 40 µl of DCS3x_A in column 11 and 40 µl DCS3x_B in column 1. The rest of the wells were filled with 20µl of B. Then 1/2 serial dilutions were performed. The same procedure was used in the fourth area preparation adding first 40 µl of DMSO3x in column 11. Plate 2 (Protein plate) contained 20 µl of P3x in rows from A to D and I to L, and 20 µl of B in rows E to H and M to P. The third plate was used for the enzyme assay and contained 10 µl of S3x in every well.

Using a Beckman Biomek FX, 10 µl from the CPD plate were transferred into the assay plate. To start the reaction, 10 µl from the Protein plate were added to the assay plate (see final plate layout in Figure 4.13). The mixture was incubated for 60 minutes at room temperature. Then, the standard BIOMOL Green procedure (section 4.2.5.1) was applied.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	No inhibitors added										DCS dose-response (high concentration)									
A											300	150	75	37.5	18.8	9.4	4.7	2.3	1.2	0.6
B											300	150	75	37.5	18.8	9.4	4.7	2.3	1.2	0.6
G											300	150	75	37.5	18.8	9.4	4.7	2.3	1.2	0.6
H											300	150	75	37.5	18.8	9.4	4.7	2.3	1.2	0.6
I	30	15	7.5	3.8	1.9	0.9	0.5	0.2	0.1	0.1	27	13.5	6.8	3.4	1.7	0.8	0.4	0.2	0.1	0.1
J	30	15	7.5	3.8	1.9	0.9	0.5	0.2	0.1	0.1	27	13.5	6.8	3.4	1.7	0.8	0.4	0.2	0.1	0.1
O	30	15	7.5	3.8	1.9	0.9	0.5	0.2	0.1	0.1	27	13.5	6.8	3.4	1.7	0.8	0.4	0.2	0.1	0.1
P	30	15	7.5	3.8	1.9	0.9	0.5	0.2	0.1	0.1	27	13.5	6.8	3.4	1.7	0.8	0.4	0.2	0.1	0.1
	DCS dose-response (low concentration)										DMSO dose-response									

Figure 4.13: The dose-response experiment plate layout. Concentrations of the compounds are annotated in each well. Thick black lines divide the four different plate areas corresponding to a positive control (no inhibitors added), dose-response for DCS at high and low concentrations, and the DMSO tolerance experiment.

4.2.6 ELF library screen

All compounds were tested at a concentration of 10 μM. Columns 1 to 4 included positive (normal reaction without compounds added) and negative controls (without enzyme). Compounds were dispensed using an Echo acoustic dispenser from Labcyte. The reaction solutions contained the same components as buffer B2x and P2x in section 4.2.5.4. The assay was allow to proceed for 1h and then 30 μl of BIOMOL Green and 5 μl of 20% (w/v) tri-sodium citrate were added as described above.

Data were first analysed using ActivityBase XE (www.idbs.com). This program tracks each well with the compound it contains and calculates the signal-to-noise ratio, the Z' score and the percentage effect (Equation 4.6) seen at every plate/well. This data can be processed using Microsoft EXCEL to select hits. A compound was considered a hit when the normalized percentage of effect was higher than the median plus three times the standard deviation.

$$\% \text{ effect} = 100 - \frac{100 \text{ signal} - \text{mean(max)}}{\text{mean(max)} - \text{mean(min)}}$$

Equation 4.6: Percentage of effect equation. **signal** is the signal in wells, **max** the maximum signal control, and **min** minimum signal control.

4.2.6.1 Validation of hits

A two step assessment of compound activity was carried out to validate the quality or usefulness of the hits.

Step 1: Replica plate and BIOMOL Green assay interference

From the ELF screening, 74 compounds were considered hits. These compounds were dispensed to a final compound concentration of 20 μM and randomly located (in duplicate) in the plate. Three compound plates were set up to perform different assays. The replica plate contained 15 μl of B2x (10 mM D-Ala, 0.4 mM ATP, 100 mM Tris-

HCl pH 8, 10 mM KCl and 10 mM $MgCl_2$) and 15 µl of P2x (0.4 nM *Bp*Ddl) except in the area A1 to L4 where 15 µl of B was placed instead of P2x. It was also necessary to check if there is any interference of the compounds with the assay. To assess this, plates containing the BIOMOL Green mixture but in the absence of the *Bp*Ddl assay components were set up. The BIOMOL–P_i plate contained the same compounds in the same order as the replica plate but only 30 µl of buffer containing 100 mM Tris-HCl pH 8, 10 mM KCl and 10 mM $MgCl_2$ was dispensed in each well. In the case of the BIOMOL+P_i plate, 30 µl of 100 mM Tris-HCl pH 8, 10 mM KCl and 10 mM $MgCl_2$ in the presence of 70 µM phosphate were added to the wells. The amount of phosphate needed in this assay had to be similar to the concentration that is released by *Bp*Ddl by the time the assay is stopped. To calculate this, a standard curve was determined from experiments at different phosphate concentrations (40, 20, 10, 5, 2.5, 1.3, 0.6 and 0 µM) in a final volume of 30 µl, and in triplicate. From the regression line equation (obtained using Microsoft EXCEL), 70 µM phosphate was determined as the approximate phosphate concentration and used on the BIOMOL+Pi plate. This validation step would indicate false positives. Data were analysed using ActivityBase XE and Microsoft EXCEL. Compounds inhibiting more than 30% of *Bp*Ddl activity were selected for the IC_{50} determination experiment.

Step 2: IC_{50} determination

The IC_{50} or potency experiment included control wells in columns 1 to 3, and was carried out in duplicate. The same assay solutions as in Step 1, in the presence and absence of 0.01% (w/v) Tween-20 and 1 mM TCEP were used. The compounds were tested at concentrations: 20, 6.7, 2.2, 0.7,0.2, 0.08 and 0.03 µM. Data were analysed using ActivityBase XE and Microsoft EXCEL.

4.2.7 Small diversity set screen

4.2.7.1 Assay development and screening

Prior to screening, a peer review of the HTP assay with DDU staff was carried out. They suggested lowering the pH to more physiological values (7.5). The activity of the enzyme at this pH was lower than at pH 8 so further HTP assay optimisation was required to improve the signal-to-noise ratio. K_m(D-Ala) and K_m(ATP) at pH 7.5 were determined as described in sections 4.2.5.2-4 but including 0.01% (w/v) Tween-20 and 1 mM TCEP in the assay buffer, and in the presence of 1 nM *Bp*Ddl. The time course experiment was performed using 3, 2 and 1 mM D-Ala and 500 μM of ATP (following the same procedure as in section 4.2.5.1).

The small DDU diversity set was used at a final concentration of 30 μM for each compound. The reaction was carried out with 1 mM D-Ala and 0.5 mM ATP, and incubated for 12 minutes. In each plate, column 23 was used for the negative control (without enzyme) and column 24 the positive control (with enzyme and in the absence of compounds). Solutions were dispensed using a Matrix Wellmate microplate dispenser (Thermo Scientific) and the absorbance measured using a PHERAstar (BMG Labtech).

Data were first analysed using ActivityBase XE, as indicated in section 4.2.6, and Microsoft EXCEL to select hits. In this case, 30% inhibition was considered a hit.

4.2.7.2 Validation of hits

Hits were validated using a potency assay. The compounds were tested at different concentrations (1/10 dilutions starting at 100 μM) in duplicate. Duplicates were set up

in independent plates with separately prepared buffer, substrates and protein stocks. The plate layout is described in Figure 4.14. The reaction and data analysis were carried out as for section 4.2.6.1.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Compound 1 Dose response										Positive control	Negative control
B	Compound 2 Dose response											
G	Compound 3 Dose response											
H	Compound 4 Dose response											

Figure 4.14: Dose-response experiment plate layout.

Data were first analysed using ActivityBase and processed using Microsoft EXCEL to select hits and calculate IC_{50} values.

4.2.8 Biolayer interferometry and a fragment library screen

Biolayer interferometry (BLI) is a real-time method for measuring ligand binding. An incident white light is guided through a glass fiber sensor where the target protein is attached. The reflected light is registered by the instrument as the interface between the optical layer and the glass fiber, and the interface between the surface and the solution. The reflected light is divided by colour channel filters into different wavelengths. The interference pattern of the reflected light is then analysed and an interferometric profile obtained (relative intensity against wavelength plot). When a compound binds to the protein, the thickness of the biolayer surface increases and the interferometric profile shifts. This difference in the wavelength, is plotted against time (Figure 4.15, BLI

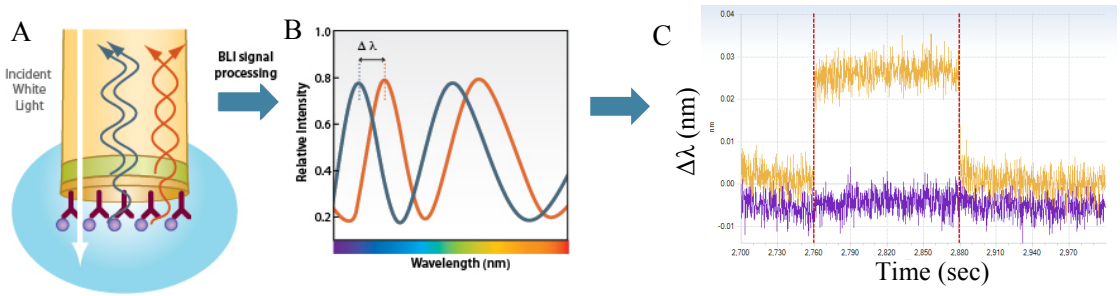


Figure 4.15: A scheme of BLI signal detection. (A) The reflected light is perceived by the instrument. (B) Interferometric profiles in the absence of compounds (blue) and in the presence of a compound that binds to the protein (orange). (C) Octet output plot. Wavelength shift (Δλ) in nm against time (seconds) plot.

applications notes at www.fortebio.com). Measurements were made with an OctetRED384 (ForteBio).

To immobilize *Bp*Ddl His_6-tag high affinity sensors were used. The surface of these sensors, like Ni-NTA columns, contains immobilised tris-nitriloacetic acid (tris-NTA) that has been charged with Ni^{2+} to assist the His_6 tail binding. The ligands were placed in 96 or 384-well plates together with reference wells (containing the assay buffer without ligands). The following protocols were used during the experiment:

1) Protein binding to the sensors: the protein is prepared in the assay buffer in a 96-well plate.
    - Column 1: 200 µl of buffer.
    - Column 2: 200 µl of protein dilution in buffer.
    - Sensor position protocol:
        1. Baseline for 60 seconds at column 1.
        2. Charge sensors at column 2 for 900 seconds.
        3. Wash at column A for 600 seconds.
2) Ligand screening 384-well plates. The sensor location program is the following:
    - Baseline: $column_{n+1}$, odd rows, for 60 seconds.
    - Association: $column_{n+1}$, even rows, for 120 seconds.
    - Dissociation: $column_{n+1}$, odd rows, for 120 seconds.
    - Baseline, association and dissociation steps for every plate column.
3) Stripping protein from the sensors:
    - Column 1: 200 µl of buffer.
    - Column 2: 200 µl of 10 mM glycine pH 1.7.
    - Column 3: 200 µl of 10 mM $NiCl_2$.
    - Sensor location protocol:

1. Baseline for 60 seconds at column 1.
2. Strip: 5 seconds at column 2.
3. Neutralize: 5 seconds at column 1.
4. Repeat 2-3 for 5 times.
5. Recharge: 60 seconds at column 3.
6. Wash: 600 seconds at column 1.

4.2.8.1 Experimental set up

Ni-NTA sensors were purchased from ForteBio, and black plates from Greiner (code 781209). Sensors used were incubated at room temperature with binding buffer (100 mM Tris-HCl pH 8, 10 mM KCl and 10 mM $MgCl_2$). A control protein (TEV protease) was used. This would inform about promiscuous binders.

Four different concentrations (0.5, 5, 10 and 20 μM) of *Bp*Ddl and TEV protease were prepared in binding buffer, and tested following protocol 1 described above. The minimum protein concentration that allows a saturation of the binding of the sensor in the experiment time was chosen for carrying out the screening.

ATP was used as a control compound. The plates used were 96-well plates, with 200 μl of buffer and ATP concentrations of 500, 166.7, 55.6, 18.5, 6.2 and 2 μM (serial 1/3 dilutions). The baseline and dissociation measurements were performed using column 1. Columns 2 to 12 contained increasing concentrations of ATP (column 12 contains the highest ATP concentration). Protocol 2 was used for this experiment.

The screen was performed using a protein concentration of 10 μM, compound concentration of 100 μM, ATP concentration of 500 μM, in a final volume of 100 μl.

Protocol 2, as described earlier, was used and protein was stripped from sensors and recharged before starting a new assay plate.

4.2.8.2 Validation of hits

From the fragment screen, 11 hits were selected (Table 4.5). As an initial validation, these hits were assayed at a concentration of 500 μM. Stock preparation of each compound is described in Table 4.5. 2 mM D-Ala was included in the assay buffer. Only one compound with an inhibition effect greater than 20% was followed up for K_i determination.

The K_i of compound #3 was determined as for DCS using final #3 concentrations of 200, 500, 2000 and 5000 μM. Stocks (10x) for every inhibitor final concentration were prepared. K_i calculation was performed in the same manner as $K_{i,DCS}$.

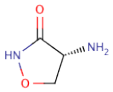
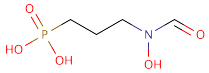
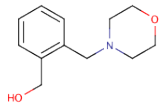
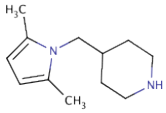
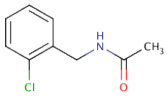
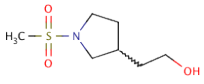
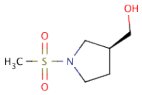
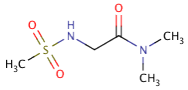
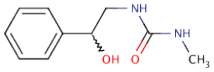
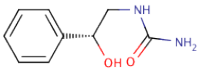
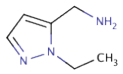
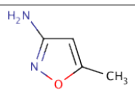
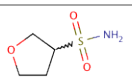
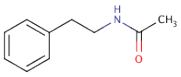
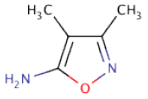
#	Name	Structure	Source	Alternative	Stock preparation
-	D-cycloserine		Sigma, C6880	-	1 M in water
-	Fosmidomycin		Molecular Probes, FR-31564	-	100 mM in water
1	[2-(Morpholin-4-ylmethyl)phenyl]methanol hydrochloride		Sigma, CBR00740	-	100 mM in water
2	4-[(2,5-dimethyl-1H-pyrrol-1-yl) methyl]piperidine		Sigma, CDS018123	-	40 mM in 10% DMSO
3	*N*-(2-chlorobenzyl) acetamide		Enamine, BBV-46868559	-	50 mM in 50% DMSO
4	2-(1-methanesulfonyl pyrrolidin-3-yl)acetic acid*		Enamine, alternative, BBV-33919693, (1-methanesulfonylpyrrolidin-3-yl)methanol		100 mM in water
5	2-methanesulfonamido-*N*,*N*-dimethylacetamide		Enamine, BBV-46887175	-	100 mM in water
6	(2-hydroxy-2-phenylethyl)urea*		Enamine, alternative, BBV-24876420, (2-hydroxy-2-phenylethyl)urea		100 mM in water
7	c-(2-ethyl-2H-pyrazol-3-yl)-methylamine		Sigma, CDS006901	-	100 mM in water
8	3-amino-5-methylisoxazole		Sigma, 232270	-	1 M in water
9	oxolane-3-sulfonamide		Enamine, EN300-75960	-	100 mM in water
10	*N*-phenethylacetaminde		Sigma, S581887	-	100 mM in water
11	5-amino-3,4-dimethylisoxazole		Sigma, 280240	-	100 mM in water

Table 4.5: Information on compounds tested as potential inhibitors. # is the number given to the fragment screening hits. The full name, structure, and suppliers are detailed in columns 2 to 4. Some compounds were not commercially available (*) so an alternative compound was purchased (structure in column 5). The last column indicates the compound stock preparation.

4.3 Results and discussion

4.3.1 Recombinant protein production

Efficient recombinant protein production systems were prepared. Yields for both proteins, *Bp*Ddl (mass 33.3 kDa) and *Ft*Ddl (mass 32.8 kDa), were about 20 mg per litre. SEC was carried out to investigate the quaternary structure of Ddl. From previous studies it is expected to observe Ddl dimerisation (Meziane-Cherif *et al.,* 2010). The results showed that the recombinant *Ft*Ddl is heterogeneous with at least three species with estimated masses 27, 75 and 137 kDa. In the case of *Bp*Ddl, single species of apparent mass of 42 kDa, has been identified (Figure 4.16). The observed values are different from the expected masses and in the case of *Bp*Ddl the difference differs by about 10 kDa. The inconsistency in the molecular masses was checked in native-PAGE

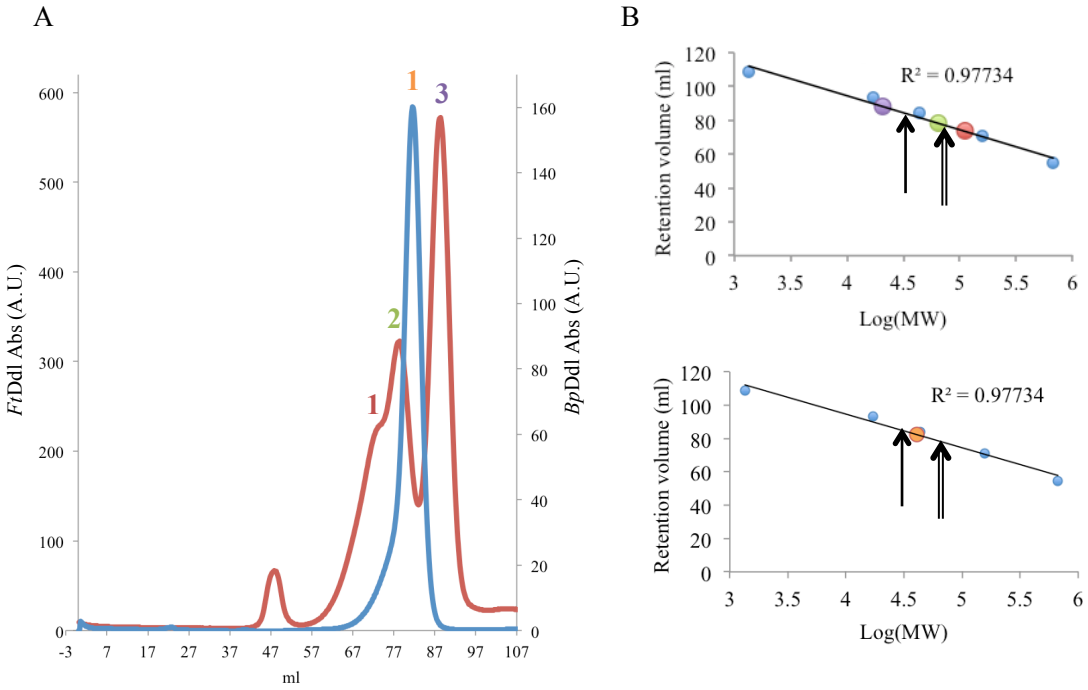


Figure 4.16: (A) SEC chromatogram for *Ft*Ddl (red) and *Bp*Ddl (blue). Ddl peaks are numbered. (B) Equilibration curves with *Ft*Ddl (top) and *Bp*Ddl (bottom) peaks plotted. Each peak is coloured differently. The expected value for monomers is marked with an arrow and dimer with a double arrow.

gels (Figure 4.17). For *Ft*Ddl, the peak corresponding with the monomer was kept and used in native-PAGE and crystallisation. Several species were found in the native-PAGE indicating there might be a dynamic mixture of quaternary structures. A very light band corresponding to the monomer can be seen for a more concentrated sample. Well defined bands can be identified as dimer (66 kDa) and hexamer (194 kDa), octamer (260 kDa) and tetradecamer (420 kDa). This is consistent with the result from the SEC. In the case of *Bp*Ddl, two bands corresponding with about 43 kDa (monomer) and 106 kDa can be seen. The second band could be an aggregation product as the samples for native-PAGE were more concentrated than the used for SEC.

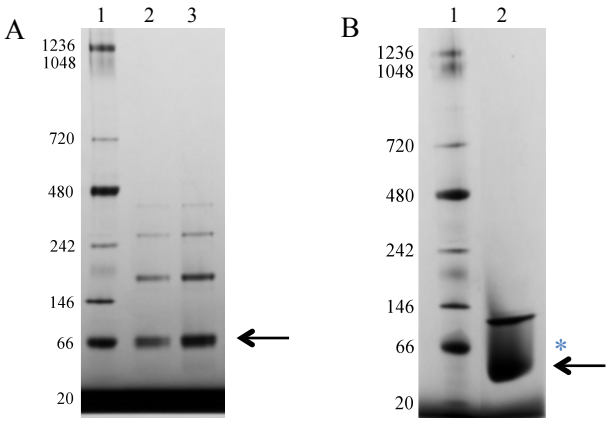


Figure 4.17: Native-PAGE 3-12%. Molecular weight standards are in lane 1. (A) *Ft*Ddl Native-PAGE with 3 ng (lane 2) and 5 ng (lane 3) of sample. The arrow indicates the dimer. (B) *Bp*Ddl Native-PAGE with 10 ng (lane 2) of sample. The arrow indicates the monomer and * the expected value for *Bp*Ddl dimer.

In native-PAGE, using a lower *Bp*Ddl concentration, two bands of about 43 kDa and 106 kDa were observed. The apparent Ddl molecular weight using SEC and native-PAGE differs from the expected values determined from the amino acid sequence. However, in SDS-PAGE, both proteins migrate with a molecular weight of about 30 kDa. The difference between the *Bp*Ddl apparent mass and the expected value could be due to presence of a cysteine on the surface of *Bp*Ddl (Cys232, not present in *Ft*Ddl) as it might cause disulfide bond formation. The quaternary structure could also be analysed using SAXS (small-angle X-ray scattering). The protein concentration used for the crystallisation trails is 20-50 times greater than the samples used for SEC and native-

PAGE. If the dimer or oligomer formation is dependent on protein concentration, this implies the samples used for crystallisation could have a different assembly/species components to those used for quaternary structure investigation. This could be further investigated with DLS (dynamic light scattering) using different sample concentrations and would indicate if the samples used for crystallisation were monodisperse or contained several species.

4.3.2 Crystallisation and structure determination

Two different *Ft*Ddl samples were screened; one corresponding with the SEC dimer peak and the other with the monomer. Small crystals were obtained with both samples (Figure 4.18) in conditions H9 from JCSG plus and G1-G2 from The Classics Suite. Optimisation plates for *Ft*Ddl were set up but did not lead to improvements in crystal size. This might be due to the presence of a non-homogeneous sample (see Figure 4.17).

For *Bp*Ddl, six hits from various JCSG plus and The Classics Suite conditions were obtained. All *Bp*Ddl crystals grew in clusters. *Bp*Ddl crystal optimisation was performed for the conditions H12 from The Classics Suite, and A2 and H9 from JCSG plus. Only needle-shaped clustered crystals grew (Figure 4.18). These needles were used for microseeding, which led to an improvement. The biggest crystals were obtained with a 1/125 diluted seed stock (Figure 4.18C) and diffracted to 2.5-2.0 Å using the in-house X-ray generator. At the Diamond Light Source beamline I04 data to beyond 2 Å were obtained. Co-crystallisation was carried out with various ligand combinations and three structures of *Bp*Ddl were completed. The crystallographic statistics are given in Table 4.6. Two contained AMP (at 2.0 Å and 1.3 Å resolution) and the third had AMP and D-alanyl-D-alanine bound (1.5 Å). Another 11 data sets were collected from crystals grown in conditions containing different ligand

combinations. Even when no co-factor was added, the structures had AMP bound. To check whether the samples contained AMP or it was a contaminant from the crystallisation components, an assay for the determination of AMP was performed (section 4.3.4.2).

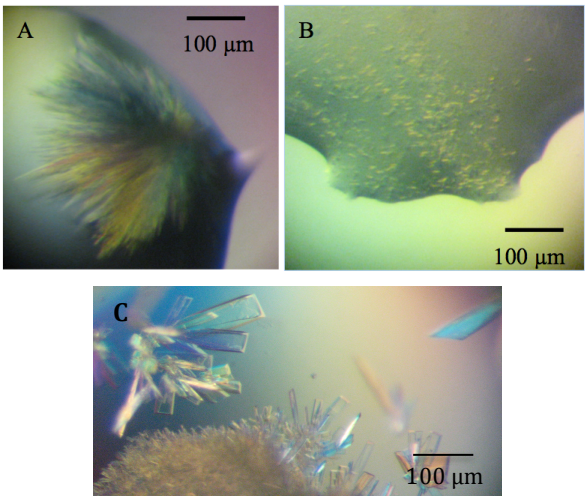


Figure 4.18: (A) *Bp*Ddl crystals. Commercial screening The Classics Suite from Qiagen, condition E9. Ratio protein: precipitant 1:1. (B) *Ft*Ddl crystals. Commercial screening JCSG Plus from Molecular Dimensions, condition H9. Ratio protein: precipitant 1:1. (C) Optimised *Bp*Ddl crystals. Condition in the drop: 0.1 M Li_2SO_4, 0.05 M Bis-Tris pH 5.5, 12.5% (w/v) PEG 3350 and 12.5 mg ml^{-1} protein.

Structure	***Bp*Ddl:AMP-1**	***Bp*Ddl:AMP-2**	***Bp*Ddl:D-Ala—D-Ala**
Space group	$P2_1$	$P2_1$	$P2_1$
Wavelength (Å)	0.979895	0.917410	0.917410
Unit cell dimensions *a*, *b*, *c* (Å), β (º)	69.58, 61.23, 70.4 90.18	69.65, 61.15, 70.08 90.17	69.64, 61.13, 69.67 90.31
Resolution range[a] (Å)	45.97 - 2.0	49.33 – 1.3	49.23 – 1.5
No. Reflections	131259	672251	283077
Unique reflections	39982	141049	90609
Completeness (%)	99.4 (99.7)	97.7 (95.7)	96.6 (98.3)
R_{merge}[b]	0.108 (0.608)	0.04 (0.665)	0.049 (0.448)
Redundancy	3.3	4.8	3.1
< I/σ(I)>	8.7 (2.1)	15.0 (2.1)	10.5 (2.1)
Wilson *B* (Å^2)	19.20	14.58	15.40
R_{work}[c]/R_{free}[d]	0.1802 / 0.2415	0.1486 / 0.1917	0.1321 / 0.2071
DPI[e] (Å)	0.20	0.05	0.08
Bond lengths (Å) / angles[f] (º)	0.0160 / 1.8103	0.0172 / 2.0453	0.0269 / 2.5848
Average *B*-factors (Å^2)	18.48	23.94	25.23
Protein residues	913	1412	1345
Water molecules	292	785	709
Metal ions	0	1 (Mg^{2+})	1 (Mg^{2+})
Ligands	8 (2 AMP, 3 SO_4^{2-}, 3 ethylene glycol)	7 (2 AMP, 2 SO_4^{2-}, 3 ethylene glycol)	15 (2 AMP, 1 D-Ala—D-Ala, 3 SO_4^{2-}, 8 ethylene glycol, 1 poly-ethylene glycol)
Ramachandran analyses			
Favoured regions (%)	97.2	98	98
Allowed regions (%)	99.8	99.8	99.8
Outliers	D5, A55, G75 and D257 (subunit B)	G153, S155 (subunit A) and D257 (subunit B)	G153 (subunit A) and D257 (subunit B)

Table 4.6: Crystallographic statistics. [a.] Values in parentheses refer to the highest resolution shell. [b.] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. [c.] $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. [d.] R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations. [e.] Diffraction Precision Index (Cruickshank, 1999). [f.](Engh and Huber,1991).

4.3.3 Ddl structure

4.3.3.1 Overall structure

Previous work showed Ddl is a member of the ATP-grasp superfamily. This family presents a characteristic and highly conserved fold despite the low sequence identity (about 10-20%) that the different proteins share. ATP-grasp proteins contain three well-defined domains and the ATP binding pocket is the region with the highest structure conservation (Fawaz *et al.,* 2011). In the case of Ddl, the domains are called N-terminal, central and C-terminal (Figure 4.19). *Bp*Ddl secondary structure comprises 10 α-helixes, one 3_{10}-helix, one π-helix and 12 β-strands (Figure 4.20). As in other Ddl proteins (Fan *et al.,* 1997; Lee *et al.,* 2006; Wu *et al.,* 2008; Bruning *et al.,* 2011; Doan *et al.,* 2014), the N-terminal domain presents an α/β-domain and the central and C-terminal domains show the ATP-grasp fold; two α/β-domains that generate a narrow pocket for the ATP binding. The active site is located in the C-terminal domain whereas residues from the central and the C-terminal domains form the ATP binding pocket. Two loops are implicated in the binding of the co-factor and the substrates. The P-loop (residues 149 to 159) is located in the central domain, between β-strands 5 and 6, and

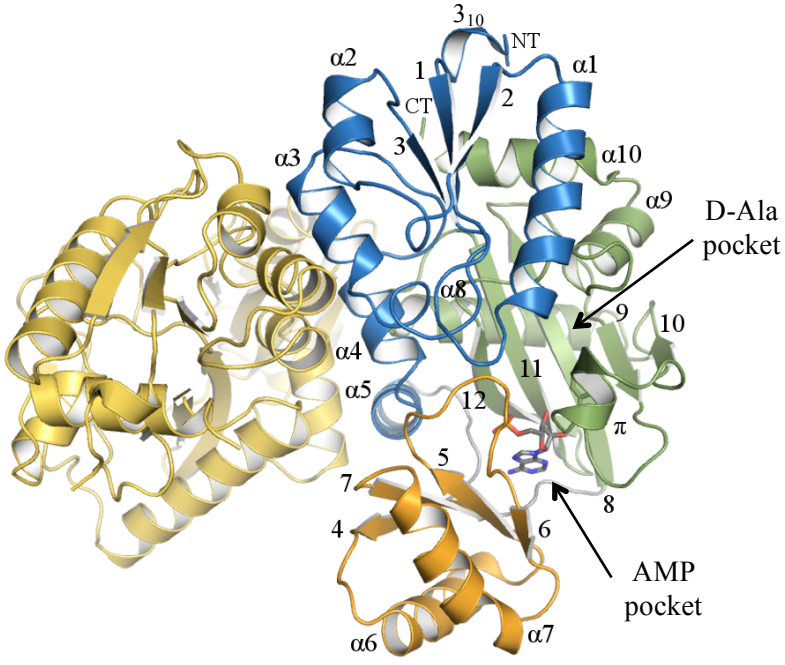


Figure 4.19: Cartoon representation of the *Bp*Ddl:AMP-2 structure. Subunit A is coloured in blue, green and orange, and subunit B in yellow. AMP is represented in sticks. α-helixes and β-strands have named as in Figure 4.20. 3_{10} and π helixes are marked. The N and C-termini are labelled as CT and NT.

interacts with the γ-β phosphates of ATP. The ω-loop is located in the C-terminal domain and presents a π-helix configuration involving residues 198 to 222.

*Bp*Ddl has a dimer in the asymmetric unit where each subunit is aligned perpendicularly to each other. Conformational differences are noted between monomers (Figure 4.21), which might correspond to structural changes that Ddl undergoes during catalysis. One dimer is present in the asymmetric unit of *Bp*Ddl:AMP-2 and *Bp*Ddl:D-Ala—D-Ala, and subunits were named A and B.

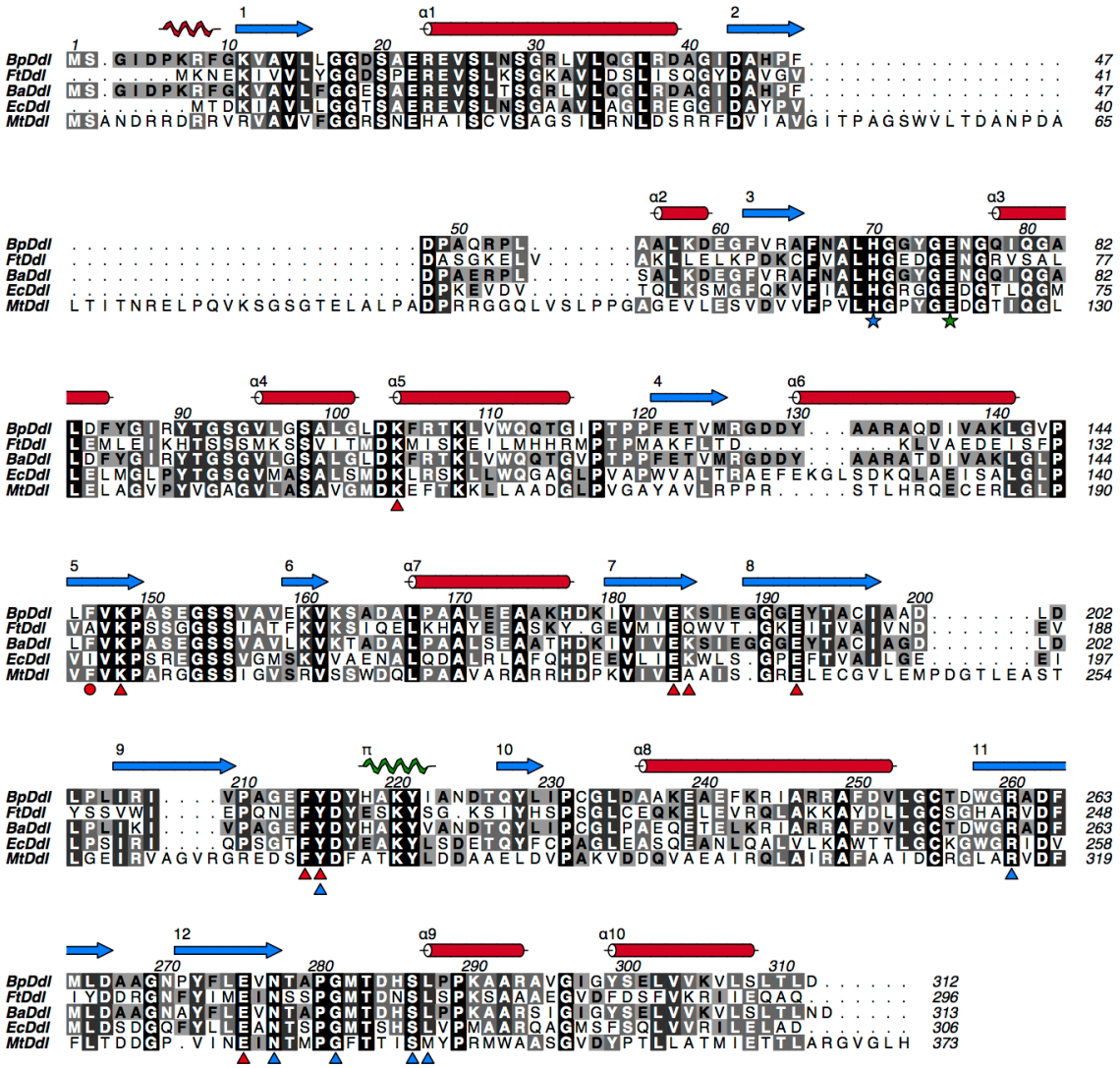


Figure 4.20: Alignment of *Bp*Ddl, *Ft*Ddl, *Ba*Ddl (*B. ambifaria* Ddl), *Ec*Ddl (*E. coli* Ddl) and *Mt*Ddl (*M. tuberculosis* Ddl). *Bp*Ddl secondary structure is included. β-strands are represented as blue arrows, α-helixes as red cylinders, 3_{10}-helix as a red spiral, π-helix as a green spiral. Amino acids that interact with ATP are marked with red triangles and the residue with stacking interactions with a red circle. Active site pocket residues are indicated with blue triangles/star and the metal binding residues as stars.

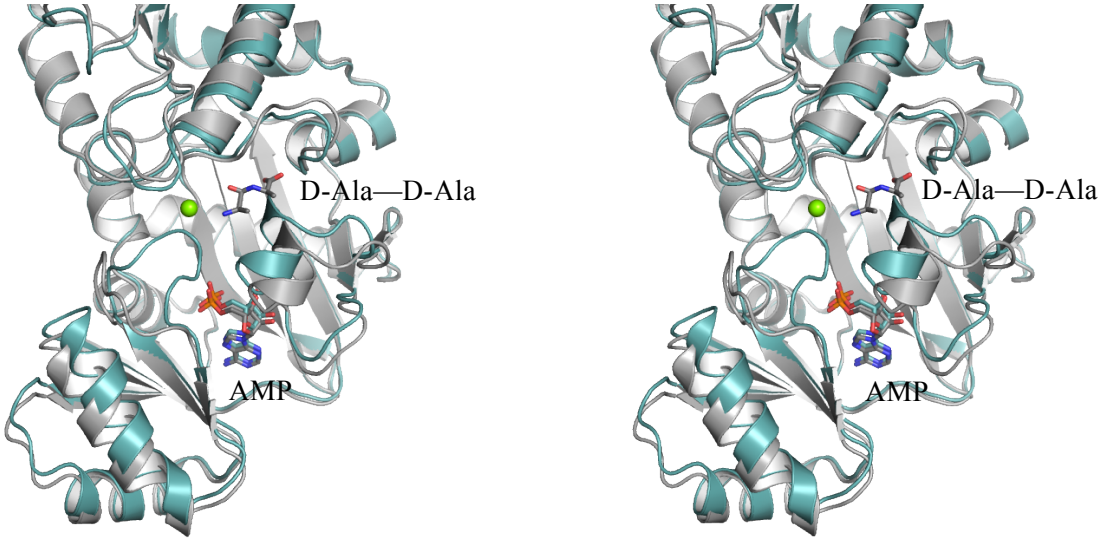


Figure 4.21: Stereo view of the C_α overlay of subunit A (in green) and B (grey) from *Bp*Ddl:D-Ala—D-Ala. The green sphere is Mg^{2+}.

Structural overlays indicate that the conformational differences between subunits are conserved in *Bp*Ddl:AMP-2 and *Bp*Ddl:D-Ala—D-Ala structures. The r.m.s.d. (root-mean-square deviation) values from the overlays are shown in Figure 4.22. This difference in conformation is not observed in other previously characterised dimeric Ddl structures (shown in Table 4.7). The *Bp*Ddl dimer is formed by about 12-13% of the monomer accessible surface area (ASA) and involves residues from α3, α4, α5 and the C-terminal region of α8. Looking at the available structures from different organisms (Table 4.7), the dimer arrangement is shared in all organisms, and the percentage of ASA involved in the dimerisation, also called buried surface area (BSA), is similar in different Ddl structures (Table 4.7). These BSA values are lower than expected for a surface with a high probability of forming assemblies (at least 5000 $Å^2$). The specificity of the interaction is determined by the P-value, which measures the probability of having similar observed solvation

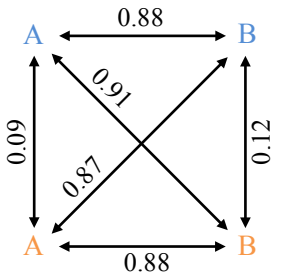


Figure 4.22: Structural comparisons. Values for r.m.s.d. (Å) for C_α overlay of *Bp*Ddl subunits. *Bp*Ddl:AMP-2 subunits are coloured in blue and *Bp*Ddl:D-Ala—D-Ala in orange.

free energy gain ($\Delta^{i}G$) in other surface areas. Thus, when the P-value is greater than 0.5 it means the surface has a lower hydrophobicity than expected and indicates the interaction is likely due to crystal packing or unspecific molecules arrangement. In contrast, a P-value lower than 0.5 indicates the hydrophobicity is higher than expected and the interaction is probably specific (Krissinel and Henrick, 2005; Krissinel and Henrick, 2007). Despite the low BSA values, the P-values indicate the dimer assembly is specific in all the structures (Table 4.7). *Mycobacterium tuberculosis* Ddl (*Mt*Ddl, Bruning *et al.,* 2011) binds in the same manner than the rest Ddl but has a lower contact area, showing that around half of analysed residues are not involved in dimer formation, and no contacts are noticed in α3 and α8. Also, the P-value is higher for *Mt*Ddl than the other structures, consistent with these observations. Ddl structures can be grouped according to the additional number of amino acids, compared to *Bp*Ddl, at the C-terminal end of α10. When extra residues are present, α10 is elongated towards the partner subunit and more contacts assist dimer formation. This is the case for *Bacillus anthracis* (*Ba*Ddl), *Streptococcus mutants* (*Sm*Ddl, Lu *et al.,* 2010), *S. aureus* (*Sa*Ddl, Liu *et al.,* 2006), *Coxiella burneii* (*Cb*Ddl, Franklin *et al.,* 2015) and *Xantomonas oryzae* (*Xo*Ddl, Doan *et al.,* 2014), where an elongated α10, carrying from 8 to 20 additional residues, is observed. The conservation, among the different orthologues, of the residues contributing to the dimer formation in *Bp*Ddl is shown in Table 4.7. The highest conserved residues, Gly77, Gly81, Val95, Ala99 and Asp103 are present in 12-13 structures of the 14 analysed and three of them are located near the binding pockets (Gly77, Val95 and Asp103). Despite differences in the rest of the residues, the dimer arrangement seems to be a recurrent feature in these Ddl crystals.

Organism	***Bp***	***Bam***	***Bx***	***Yp***	***Ec***	***Tt***	***Tc***
PDB	***Bp*Ddl-AMP**	**4EG0**	**4EGJ**	**3V4Z**	**4C5C**	**2YZG**	**2FB9**
ID (%)	-	92	89	49	48	31	31
r.m.s.d. (Å)	0.12*	0.39	0.63	1.28	1.92	1.84	1.85
Res_{Dimer} (%)	-	100	100	45 (21)	45 (28)	41 (17)	41 (17)
ASA_A (Å^2)	13896	12529	12231	13489	13186	14918	15331
BSA_A (Å^2)	1793	1287	1089	1331	1666	1372	1368
BSA_A (%)	13	10	9	10	13	9	9
P-$value_A$	0.007	0.060	0.090	0.022	0.008	0.018	0.044
ASA_B (Å^2)	13163	12823	11431	13601	13109	13983	15331
BSA_B (Å^2)	1540	1309	1089	1061	1496	1372	1368
BSA_B (%)	12	10	10	8	11	10	9
P-$value_B$	0.014	0.064	0.094	0.008	0.010	0.018	0.048

Organism	***Xo***	***Ba***	***Sa***	***Cb***	***Sm***	***Mt***	***Hp***
PDB	**4ME6**	**3R23**	**2I87**	**3TQT**	**3K3P**	**3LWB**	**2PVP**
ID (%)	30	34	25	24	28	30	23
r.m.s.d. (Å)	1.54	1.60	1.73	1.95	2.33	1.63	3.76
Res_{Dimer} (%)	38 (21)	24 (28)	24 (28)	24 (24)	24 (21)	17 (14)	17 (34)
ASA_A (Å^2)	16009	14940	16632	16012	14731	14807	16997
BSA_A (Å^2)	1800	1433	2329	2294	1557	954	1454
BSA_A (%)	11	10	14	14	11	6	9
P-$value_A$	0.013	0.046	0.176	0.021	0.169	0.382	0.158
ASA_B (Å^2)	16041	14842	16879	15532	14669	13426	16365
BSA_B (Å^2)	1800	1281	2325	2294	1557	959	1454
BSA_B (%)	11	9	14	15	11	7	9
P-$value_B$	0.011	0.041	0.083	0.003	0.187	0.333	0.091

Table 4.7: Comparison of *Bp*Ddl with orthologues. The first section of the table shows, the percentage of the protein sequence similarity (ID), r.m.s.d. values of the C_α overlay of *Bp*Ddl-AMP-2 subunit B and a selected Ddl monomer, and the percentage of resides in the dimer interface (Res_{Dimer}) that are conserved and semi-conserved (in brackets). The second part includes the ASA, BSA values for each subunit (indicated in the subscript), and P-values (probability of the specificity of the interface). BSA is noted as Å^2 and %. (*) overlay performed between subunits B from *Bp*Ddl:AMP-2 and *Bp*Ddl-D-Ala—D-Ala. *B. anthracis* (*Ba*), *B. ambifaria* (*Bam*), *B. pseudomallei* (*Bp*), *B. xenovorans* (*Bx*), *C. burneii* (*Cb*), *E. coli* (*Ec*), *H. pylori* (*Hp*), *M. tuberculosis* (*Mt*), *S. aureus* (*Sa*), *S. mutants* (*Sm*), *Thermus thermophilus* (*Tt*), *T. caldophilus* (*Tc*), *X. oryzae* (*Xo*) and *Y. pestis* (*Yp*).

Comparison of the *Bp*Ddl structures with other available structures showed the highest homology with the *B. ambifaria* Ddl (92% identity and 0.6 Å r.m.s.d. for the monomer C_α overlay). As expected, other proteins from the ATP-grasp superfamily, such as acetyl-CoA-carboxylase and biotin carboxylase (18-19% identity and 2.8-2.7 Å r.m.s.d.), were also noted as structural homologues. The major structural similarity in the ATP-grasp fold family is found in the central and C-terminal domains. A comparison of *Bp*Ddl:AMP-2 structure with a representative of the ATP-grasp superfamily (Figure 4.23) shows the high structural homology that this superfamily presents especially in the ATP-binding site. Major structural differences were found at the N-terminal domain and the position of the central domain (more open or closed conformations).

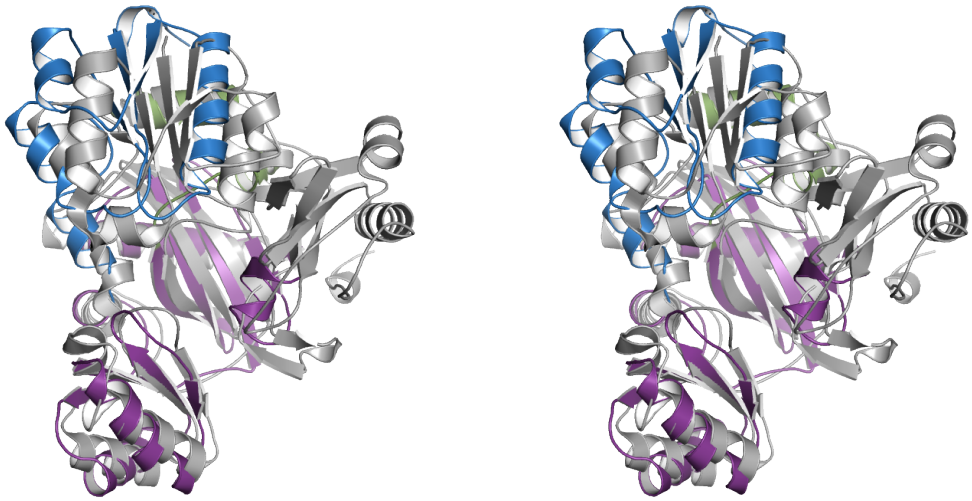


Figure 4.23: Stereo view of the overlay of *Bp*Ddl:AMP-2 and glycinamide ribonucleotide synthetase (PurD, Sampei *et al.,* 2010, PDB code 2YW2), two members of the ATP-grasp superfamily. The r.m.s.d. value for the overlay is 4.56 Å. The N-terminal domain is coloured in blue, the central domain in green and the ATP-grasp domain in purple. PurD is coloured in grey.

4.3.3.2 Co-factor and substrate pockets

Both co-factor and substrate binding sites are shown in Figure 4.24. They are highly conserved in Ddl sequences (Figure 4.20). Indeed, the ATP-binding site residues (Lys104, Gly153, Glu184, Asp262, Glu275 and Asn277) are also conserved within

different functional proteins from the ATP-grasp superfamily (Fawaz *et al.,* 2011). The recognition of ATP involves residues from the central and C-terminal domains. These include Lys104, Lys148 and Glu275 binding to the phosphate, and Glu184, Lys185, Ile187, Glu192, Phe214 and Tyr215 that are involved in adenosine recognition. In the case of D-Ala—D-Ala, the residues interacting are Tyr215, Tyr221, Asp277, Arg260, Gly281, Ser286, Leu287 and a Mg^{2+}-bound water molecule. The D-Ala—D-Ala peptide is poorly defined at the N-terminus and was refined with zero occupancy.

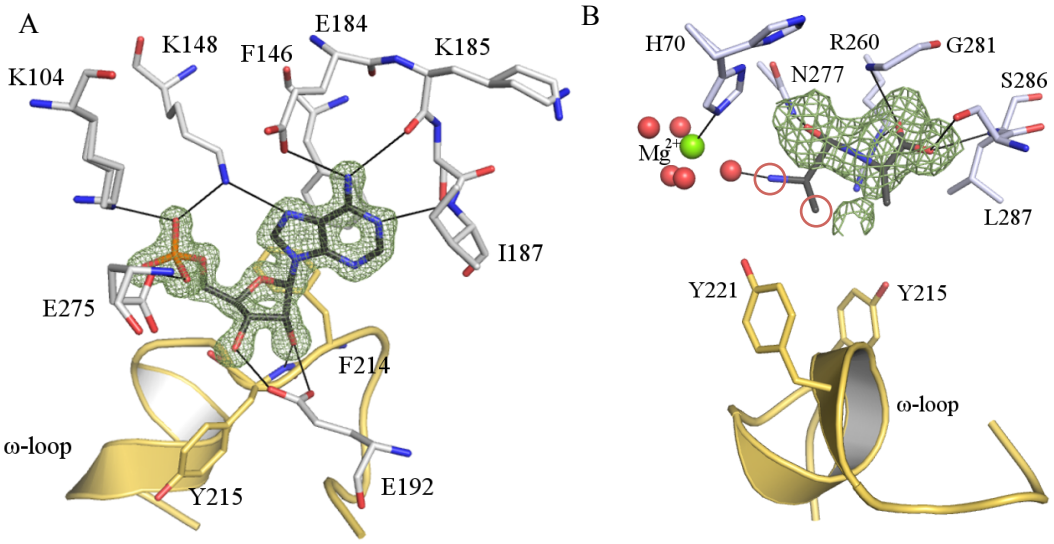


Figure 4.24: (A) ATP binding pocket. AMP is represented as dark grey sticks and the 2Fo-Fc map (5.0 σ) is shown as a green mesh. (B) D-Ala binding pocket. The dark grey sticks represent D-Ala—D-Ala. Red circles indicate atoms with zero occupancy. The ω-loop is represented as a yellow cartoon. The Fo-Fc map is represented as a green mesh (2.3 σ). Hydrogen bond interactions are indicated with black lines.

An alignment of the protein sequences available in UniProt showed high conservation in residues from both pockets. In the D-Ala pocket, residues Lys220, Arg260, Asp277 and Gly281 are conserved in all sequences and Leu287 in 90%. These are key residues for the catalysis according to the proposed reaction mechanism as they polarise and keep the substrates and the intermediate in place (Figure 4.2). Less conservation is noticed for Phe214, Tyr215 and Tyr221 (40-50%) but other aromatic residues are in these positions. These amino acids could be involved in placing ATP and D-Ala in the correct position for catalysis. In the ATP pocket, Lys104, Lys148, Glu184, Glu192 and

Glu275 are conserved in 98% to 100% of the sequences. These residues are likely to be important for ATP binding.

When *Bp*Ddl:D-Ala—D-Ala was compared with other product and substrate-bound structures (from *Ec*Ddl: 1IOV, 4C5B, 4C5C; from *Tt*Ddl: 2ZDH and 2ZDQ, including ADP+phosphinate, ADP+D-Ala, ATP+D-Ala—D-Ala, ADP+D-Ala, ATP+D-Ala respectively), differences in the substrate and product binding were found (Figure 4.25). The C_α overlay of *Bp*Ddl:D-Ala—D-Ala (subunit A) with these ligand-bound structures showed the position of D-Ala—D-Ala is different from previously characterised proteins. The product in *Bp*Ddl:D-Ala—D-Ala is located further from the co-factor. This is highlighted when overlaying the ATP and AMP molecules instead of the C_α (Figure 4.25B). Additionally, the position of the Mg^{2+} is different. In other Ddl structures (*e.g. Ec*Ddl, PDB code 4C5C) two Mg^{2+} ions interact with Asp262, Glu275 and γ-β phosphates from ATP. However in *Bp*Ddl:D-Ala—D-Ala only one Mg^{2+} is observed, interacting with His70. The overall *Bp*Ddl:D-Ala—D-Ala conformation presents a more open form than in other structures, suggesting it is an intermediate

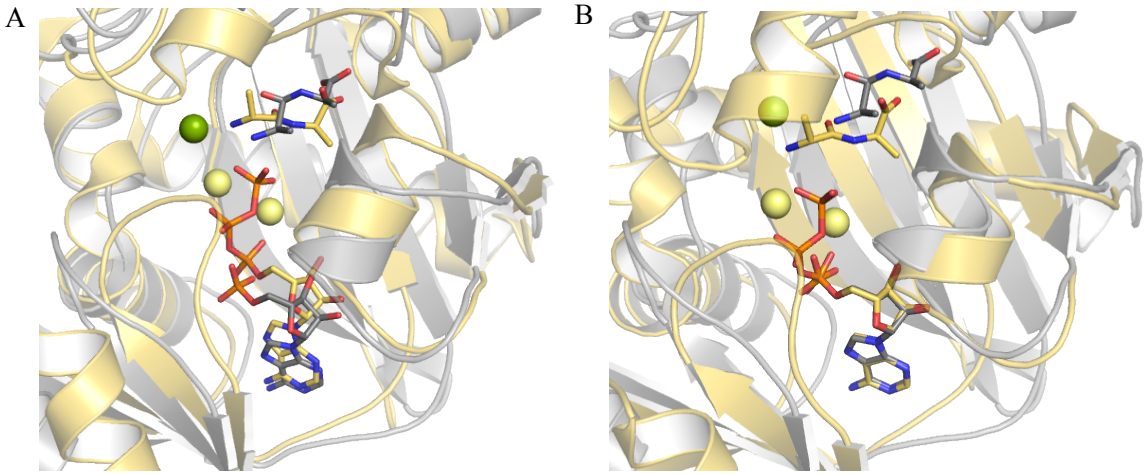


Figure 4.25: Overlay of *Bp*Ddl:D-Ala—D-Ala (grey) and *Ec*Ddl (45C5, in yellow). Ligands are represented as sticks. Mg^{2+} are represented as spheres coloured in green (*Bp*Ddl:D-Ala—D-Ala) or yellow (*Ec*Ddl). (A) C_α overall overlay. (B) Overlay using the co-factors.

conformation. However, it cannot be distinguished if this corresponds to a substrate binding or product release phase.

4.3.3.3 Conformational change

In common with other Ddl examples, *Bp*Ddl undergoes a large conformational change, involved in configuring the active site, when the co-factor binds (Figure 4.26). Binding generates a more compressed conformation where the central domain moves towards the centre of the structure and the ω-loop rotates and generates the active site pocket.

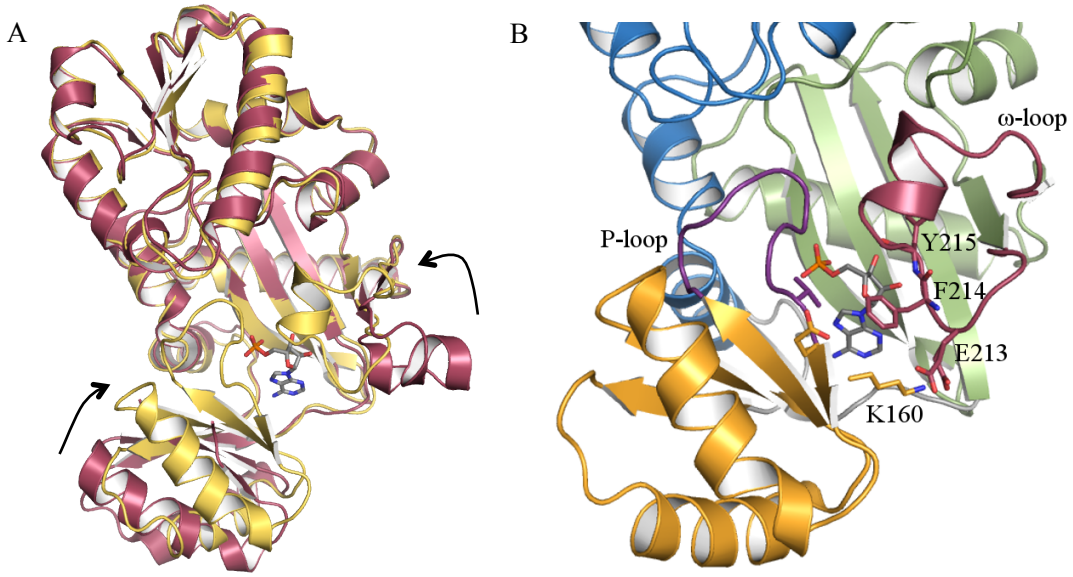


Figure 4.26: (A) Ribbon diagram of the superposition of the apo-*Bp*Ddl (4EGQ in dark red) and *Bp*Ddl:AMP-2 (yellow). AMP is represented with sticks. (B) Residues implicated in co-factor binding. Blue indicates the N-terminal domain, orange the central domain and green the C-terminal domain.

ATP binding stabilises the ω-loop conformation enforcing interactions with Tyr215 and helps to create the closed confirmation. Phe214 is able to interact with a hydrophobic area of the central domain and Glu213 with Lys160 to close the ATP binding pocket. The conformational change of the central domain involves a movement/shift of α7 of about 12.5 Å and 7.6 Å for α6 (average distances from both subunits) favoured by the AMP and ω-loop interactions. Additionally, the conformation of the P-loop is stabilised

by AMP interactions, the ω-loop and the N-terminal domain residues (Figure 4.21). In apo-*Bp*Ddl, the P-loop is disordered and not included in the model but in the AMP-bound structures this loop is ordered with an average *B*-factor of 36.7 $Å^2$ (the overall average *B*-factor value is 24.6 $Å^2$).

There are apo and ligand-bound structures for *Sa*Ddl (2I8C, 2I80, apo: 2I87; Liu *et al.,* 2006) and *Tt*Ddl (2YZM, 2YZN, 2ZDH, 2ZDG, 2ZDQ, apo; 2YZG). In both cases, the conformational changes are not as large as in *Bp*Ddl. Movement of the *Tt*Ddl central domains involves an α7 translation of 4.7 Å, while no differences in *Sa*Ddl are observed. As indicated in section 4.2.3 and 4.3.4.2, purified *Bp*Ddl contained AMP and apo-*Bp*Ddl could not obtained. With this in mind, density maps of the apo *Sa*Ddl and *Tt*Ddl structures were inspected to check for missed co-factors. Both *Tt*Ddl and *Sa*Ddl showed no evidence of co-factors in the ATP pockets. This might suggest either these apo forms are crystallographic artefacts (the conformation is driven by the crystallisation conditions), or the conformational changes are distinct among different organisms.

A comparison of the electrostatic surfaces of apo-*Bp*Ddl and *Bp*Ddl:AMP-2 indicated a difference when the co-factor is bound (Figure 4.27). The D-Ala pocket is more electropositive in the AMP-bound structure. Therefore, at pH 7.5, D-Ala (pI = 6.01) would preferably bind to the *Bp*Ddl:AMP-2 conformation over the apo form. This suggests that the D-Ala pocket is formed after the ω-loop conformational change. The residues involved in the electrostatic properties of the different conformations (Glu22, Asp262, Glu275 and Asn277 in apo-*Bp*Ddl, and His70, Arg260, and Asn277 in *Bp*Ddl:AMP-2) are highly conserved. Also, as described previously, His70, Arg260, and Asn277 are key amino acids for Ddl catalysis. This is consistent with the hypothesis

that the generation of a favourable environment for catalysis is dependent on ATP binding, and extends to the sequential ordered mechanism previously characterised (Fan *et al.,* 1994) where ATP is the first component to bind Ddl.

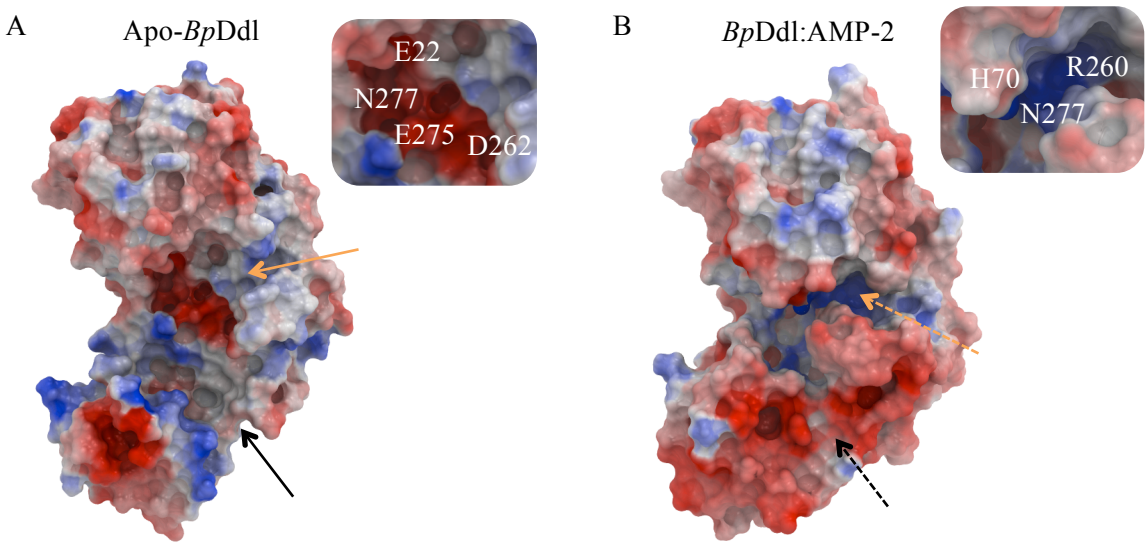


Figure 4.27: Electrostatic surface representation of (A) apo-*Bp*Ddl (4EGQ) and (B) *Bp*Ddl:AMP-2 structures. Orange arrows indicate the D-Ala pocket location and black arrows the ATP pocket. Dashed arrows indicate ATP pocket is closed and D-Ala pocket is formed. On the right side of each surface an enlargement of the D-Ala pockets are shown and amino acids marked.

4.3.3.4 Druggability

Analysis of the *Bp*Ddl:AMP-2 and the apo-*Bp*Ddl structures with the pocket finder tool from Molsoft ICM showed distinct pockets in the D-Ala and ATP binding sites (Figure 4.28). This indicates that the conformational change generates differences in the pockets. The software reports a druggability score for each pocket. A pocket is considered druggable-like when the score is greater than 0.5. The control proteins DHFR and neuraminidase showed druggability scores of 0.8 and -0.18. The values obtained for the ATP and D-Ala pockets in apo-*Bp*Ddl are -1.45 and -1.56 respectively. In the case of *Bp*Ddl:AMP-2 subunit A, these values are 0.02 and -1.60 respectively. On the other hand, *Bp*Ddl:AMP-2 subunit B presents druggability scores of 0.79 (ATP pocket) and 0.77 (D-Ala pocket). This suggests druggable-like pockets are only found in the more closed conformation.

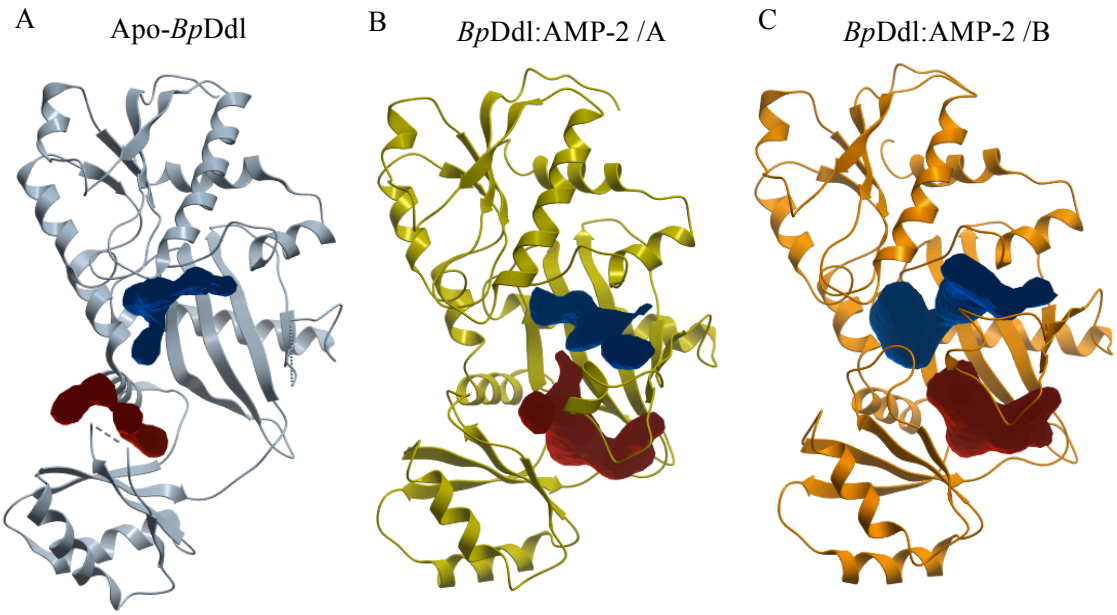


Figure 4.28: *Bp*Ddl pockets. Blue pocket is in the D-Ala binding site and red pocket in the ATP binding site. (A) Open conformation, Apo-*Bp*Ddl (4EGQ). (B) Closed conformation *Bp*Ddl:AMP-2 subunit A. (C) *Bp*Ddl:AMP-2 subunit B.

Examples of human proteins from the ATP-grasp family are glutathione synthetase (Dinescu *et al.,* 2004), inositol phosphate kinase (Josefsen *et al.,* 2007), ATP-citrate lyase (Sun *et al.,* 2011), pyruvate phosphate dikinase (Wu *et al.,* 2013), and succinyl-CoA ligase (Jaberi *et al.,* 2013). Also, other human kinases, such as the Ser/Thr kinase c-Jun N-terminal (JNK1, PDB code 4QTD, Chaikuad *et al.,* 2014), share similarities with the ATP-grasp domain. Due to the high similarity of the ATP binding pocket between different functional proteins, targeting this pocket might result in inhibition of human enzymes and trigger side effects. To avoid this, the inhibitor search was focused on obtaining new inhibitors targeting the D-Ala binding pocket.

4.3.4 Biochemical characterisation

4.3.4.1 Enzyme kinetics

Enzyme activity was investigated using a coupled spectrophotometric assay and a BIOMOL Green assay at pH 7.5 and 8. The results are shown in Table 4.8. Calculation

of K_m values were carried out using different equations (see Experimental procedures). Ddl binds to two substrates and one co-factor, and, consequently, there will be three K_m values; one corresponding to the first D-Ala (K_{m1}), the second D-Ala (K_{m2}) and the third ATP (K_{ATP}).

Using the coupled assay, inhibition of the enzyme activity was observed at high concentrations of ATP and $K_{i,ATP}$ was also determined. In the case of the BIOMOL Green assay, high ATP concentrations decrease the signal-to-noise ratio due to non-enzymatic ATP hydrolysis. This compromises the robustness of the assay and consequently, concentrations no greater than 1 mM were tested initially. Using 1 mM of ATP led to a decrease in the signal-to-noise ratio. A test at high ATP concentrations was performed and showed the same enzyme inhibition seen at the coupled assay.

For the calculations of D-Ala K_m in the BIOMOL Green assay, only K_{m2} was taken into account and data were analysed with the Michaelis-Menten equation. As high ATP concentration cannot be used with BIOMOL Green, enzyme inhibition was not observed during the experiment and K_{ATP} was determined using the Michaelis-Menten equation.

The values for K_{m2} were consistent between the coupled and the BIOMOL assays. Comparing the K_m values with other previously described Ddl enzymes, there is more similarity between *Bp*Ddl and *Mt*Ddl or *Ec*Ddl than *Hp*Ddl. Indeed, *Hp*Ddl seems to be the least related Ddl, not only having a lower amino acid percentage identity but also being a non-saturable enzyme with different kinetic parameters. For many enzymes, the K_m values are similar to the physiological concentration of the substrate (Copeland, 2010). In D-Ala producing bacteria, the physiological concentration inside the cell is

not known but it is produced in mM concentrations (Lam *et al.,* 2009) so this would match with the K_{m2} values observed.

Detection method	pH	K_{m1} (mM)	K_{m2} (mM)	K_{ATP} (mM)	$K_{i,ATP}$ (mM)	Vmax (nmol NAD^+ min^{-1})
Coupled	8	0.37 ± 0.085	4.19 ± 0.71	0.07 ± 0.01	3.16 ± 0.59	21.85 ± 0.98
Coupled	7.5	0.35 ± 0.03	5.12 ± 1.09	-	-	12.92 ± 1.38
Biomol	8	-	4.9	0.13	-	-
$Biomol_{TCEP/Tween}$	8^A	-	3.2	0.2	-	-
$Biomol_{200\mu M\ ATP}$	8^B	-	2.3	-	-	-
$Biomol_{TCEP/Tween}$	7.5^A	-	5.4	0.27	-	-
*Hp*Ddl[1]	8	1.89	627^C	0.000087	-	-
*Ec*Ddl[2]	7.8	0.0012	1.13	0.049	-	-
*Mt*Ddl[3]	8	0.075 ± 0.01	3.6 ± 0.6	0.31 ± 0.05	-	-

Table 4.8: Ddl kinetic parameters. It shows K_m values for two D-Ala and ATP, the K_i for ATP, and the V_{max} of the reaction. (^{A}) In the presence of 1mM TCEP and 0.01% (w/v) Tween-20. (^{B}) K_{m2} in the presence of 200 µM ATP instead of 500 µM ATP. (^{C}) Maximum D-Ala concentration tested. (-) Not determined. [1](Wu *et al.,* 2008), [2](Ellsworth *et al.,* 1996), [3](Prosser and de Caravalho, 2013).

The *Bp*Ddl samples used for performing the assays had AMP-bound, so the K_{ATP} determination is compromised. No information about the presence of AMP or co-factor analogues in *Hp*Ddl, *Ec*Ddl and *Mt*Ddl has been previously reported. Purification of *Mt*Ddl (Prosser and de Caravalho, 2013) and *Hp*Ddl (Wu *et al.,* 2008) was carried out using nickel-affinity columns while *Ec*Ddl used ion exchange (Zawdzake *et al.,* 1991; Ellsworth *et al.,* 1996). It might be expected that *Mt*Ddl and *Hp*Ddl also co-purified with the co-factor as the purification is similar to *Bp*Ddl, however direct confirmation should be carried out.

A test at different pH values indicated the optimum for *Bp*Ddl is pH 8. Differences in the activity due to the change in the pH are visible in the K_{m2} and the V_{max}. At a lower pH there is an increase of 1.2 times in the K_{m2} and 1.7 times decrease in the V_{max}. To further explore this, activity was measured over a pH range of 5.1 to 9.5. Regarding this, a test at different pH values has been carried out (Figure 4.29). The ATP and D-Ala

concentrations were kept constant at 10 mM D-Ala and 0.5 mM ATP. The peak of the maximum activity is observed at pH 8 (100% relative activity) and the activity decreases around 1.7 times at pH 7.5 (56% relative activity). This is consistent with the results from the K_m experiments.

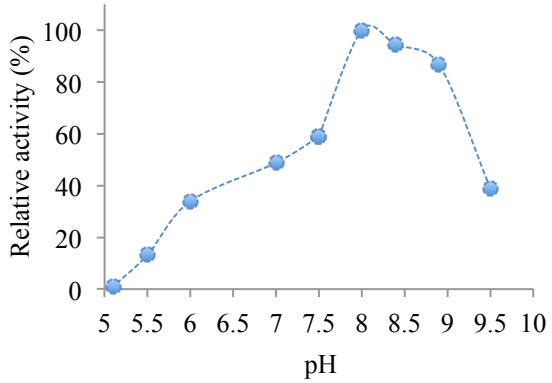


Figure 4.29: pH effect on the reaction velocity. The activity is shown as the relative activity referred to the maximum activity at pH 8.

To identify a control inhibitor to support the screening experiments, the already known Ddl inhibitor DCS and fosmidomycin, an intermediate-like compound, were tested and $K_{i,DCS}$ calculated. In the case of fosmidomycin, no inhibition was observed. Determination of the K_i and the mechanism of inhibition for DCS were assessed. Information about the mode of action of the inhibitor can be deduced from V_{max} and K_m values. If it is a competitive inhibitor, at high concentrations of substrate, the effect of the inhibitor will be nullified so the V_{max} will be the same as the non-inhibited protein. However, the binding of the substrate will be altered and the K_m will change. On the other hand, an uncompetitive/allosteric inhibitor generally inhibits by binding a distinct pocket from where substrates /co-factors bind. This will cause a decrease in the V_{max} and K_m (and increase in the specific activity K_m/V_{max}). This can be checked using a double reciprocal plot. From Equation 4.4, the intercept of the regression lines with the y axis gives the $1/V_{max}$ value. In the case of DCS, all but one of the regression lines present the same intercept (Figure 4.30). This indicates the V_{max} has not changed and suggests DCS is a D-Ala competitive inhibitor with a $K_{i,DCS}$ value of 100 μM. At the highest DCS concentration tested, a different V_{max} was observed. This is a characteristic of uncompetitive inhibitors and may indicate that the dynamics of

binding changes. DCS could be binding both D-Ala pockets and so display different binding behaviour. For *Mt*Ddl ($K_{i,DSC}$ of 25 µM) and *Ec*Ddl ($K_{i,CDS}$ of 185 µM), it was reported that slow binding of DCS occurs (Prosser and de Carvalho, 2013). Additionally, they found the slow binding is not present at high DCS concentration (1000 µM). This indicates there is a difference in binding at this concentration. A similar observation might apply to *Bp*Ddl.

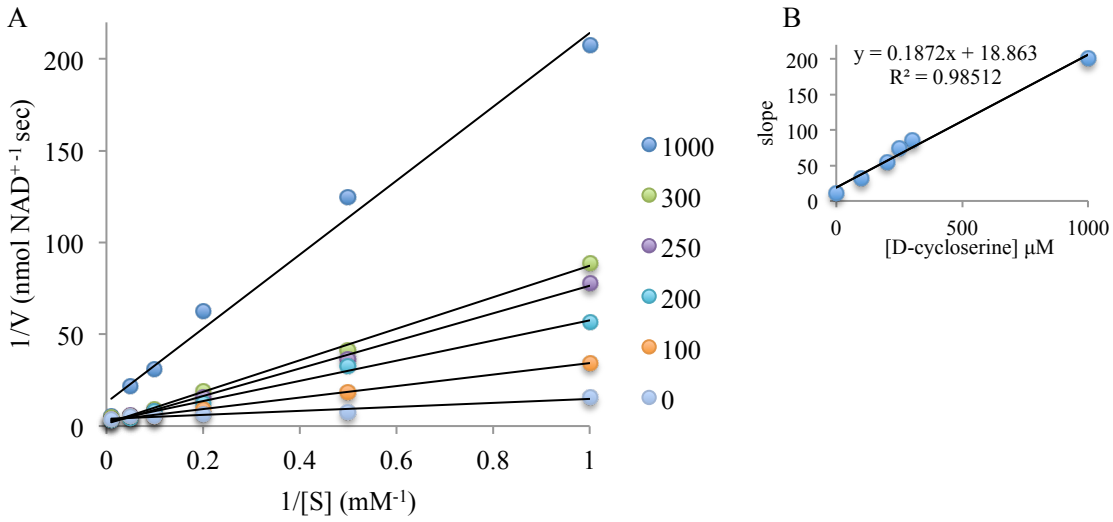


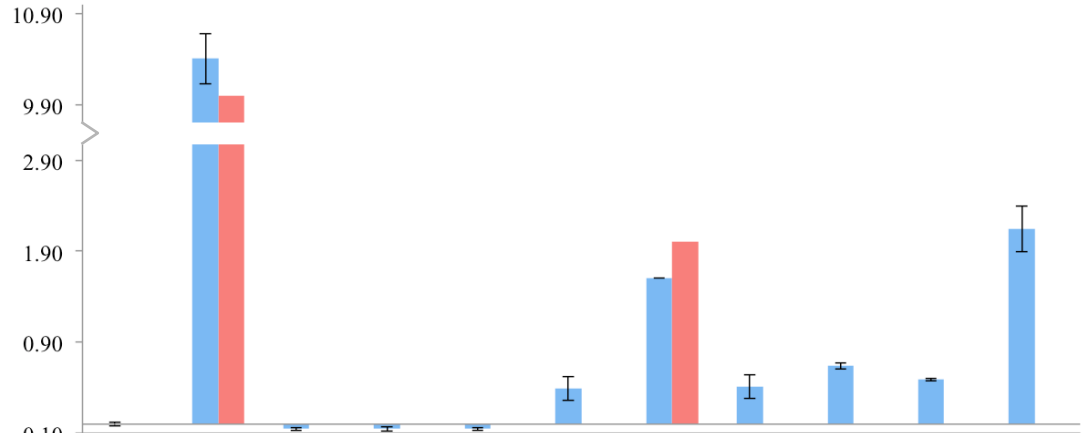
Figure 4.30: (A) Double reciprocal plot for the DCS inhibition. (B) Plot of the slopes from the reciprocal plot against the concentration of DCS. From the equation of the slope of this line the $K_{i,DCS}$ can be determined (Equation 3).

4.3.4.2 AMP detection

Characterisation of structures from a range of different conditions revealed all had AMP bound. To determine whether the AMP was present in the purified sample or not, an AMP detection assay was carried. Denatured and folded samples were compared. The AMP-Glo assay contains various enzymes, so denaturation of the samples should not interfere with assay performance.

Firstly, *Ba*KynB, *Ba*KynB in the presence of 2 µM AMP and *Bp*Ddl were treated for denaturation by 1) pH variation, 2) temperature variation and 3) chymotrypsin

digestion. The experimental and expected results are shown in Figure 4.31. The negative control (in the absence of protein and AMP) and positive control (10 µM AMP in the absence of protein) results were similar to the expected values. The *Ba*KynB sample was the same that was used for the structural analysis and confirmed there is no AMP bound (negative control). The *Ba*KynB controls inform about anomalies in the sample preparation and assay performance such as chymotrypsin interfering with the assay or signal when no AMP was present in the protein sample. The results show the *Bp*Ddl sample has been purified with AMP bound and there is an AMP signal even without pre-treatment. This indicates there is some free AMP in the protein solution (around 0.4 µM) probably due to AMP exchange/release. The *Bp*Ddl-Ch sample presented an AMP concentration of about 2.1 µM that indicates 1.6 µM (14.8 nmol) AMP belongs to AMP bound to *Bp*Ddl. Taking into account that the concentration of *Bp*Ddl is 25 mg ml^{-1} (18.9 nmol in each well), the ratio AMP:*Bp*Ddl is close to 1. This is consistent with the structural data that shows one AMP per *Bp*Ddl subunit.



	Neg	Pos	KynB WT	KynB ΔpH	KynB ΔT	KynB Ch	KynB Ch+	Ddl WT	Ddl ΔpH	Ddl ΔT	Ddl Ch
Results	0.00	10.41	-0.05	-0.05	-0.05	0.39	1.61	0.41	0.64	0.49	2.14
Expected	0	10	0	0	0	0	2	0	0	0	0

Figure 4.31: AMP detection result. **Neg** is the negative control (without enzyme, without AMP). **Pos** is the positive control from the standard curve (without enzyme with 10 µM AMP). **KynB** is *Ba*KynB sample. **WT** means no pre-treatment is performed. **ΔT** and **ΔpH** indicates variations in the temperature or pH were carried out. **Ch** indicates digestion with chymotrypsin was used and + is used when 2 µM AMP was added. The table indicates the values of the bars and the legend.

The presence of AMP in the purified enzyme has implications for the kinetic data, compound screens and crystallisation. The K_{ATP} determination and the assessment of the screening has been done with *Bp*Ddl:AMP instead of apo-*Bp*Ddl. Additionally, during the spectrophotometric assay, a retarded phase was observed during the first seconds of measurement (Figure 4.32). This could be due to the presence of AMP in the sample. Also, the fragment screen using BLI was carried out with the closed *Bp*Ddl conformation and might explain why a low ATP binding signal was measured.

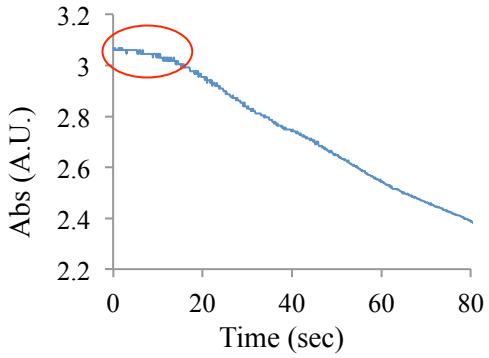


Figure 4.32: Spectrophotometric assay. Absorbance at 340 nm against time (sec) plot. The circle indicates the retarded phase of the reaction.

A *Bp*Ddl structure of the apo-form (PDB code 4EGQ) has been determined previously by the Seattle Structural Genomics Center for Infectious Disease (SSGCID). The ATP binding site was inspected to confirm it is indeed an apo-*Bp*Ddl. There was no evidence, in terms of electron density or difference peaks, to suggest the presence of the co-factor. The apo-crystal was grown in a condition of a basic pH (0.1 M Ches pH 9.5, 1 M Na-K tartrate and 0.2 M Li_2SO_4) while the crystals obtained in this work were obtained at pH 5.5. It is unclear why in our structures AMP is present yet in the PDB entry 4EGQ it is absent. However, this could be due to the different crystallisation conditions.

4.3.5 BLI: Fragment library screen

The DDU fragment library, around 700 compounds, was tested using a BLI binding assay. A control compound is required to check the sensor/protein integrity. In this case, D-Ala or DCS stripped the protein from the sensors, and background and binding

signals decreased with increasing concentrations of the compounds. This might be due to competition of D-Ala/DCS for the bound Ni^{2+}. Additionally, according to the manufacturer, these compounds are below the molecular weight edge of detection of the Octet (150 Da, BLI applications notes at www.fortebio.com). ATP was tested as a possible control compound. The binding signal was lower that expected and only at high concentrations of ATP was significant binding recorded. This can be explained by the fact that the sample contains AMP as a contaminant from the *E. coli* recombinant expression system. Fortunately, ATP also binds to TEV protease and provided an additional control.

Once the screen was carried out, the hits were selected. A compound was considered a hit when a binding response was obtained for *Bp*Ddl sensors but not for TEV protease. The hits, together with their binding responses, are shown in Table 4.9. From the 11 selected hits two gave a negative binding response, which could indicate a large conformational change, unfolding of the protein or aggregation. The 11 hits were tested at 500 μM for inhibitory properties. Only compound #3 gave some inhibitory effect. A similar K_i determination experiment to $K_{i,DCS}$ was carried out. After analysing the data, the $K_{i,\#3}$ was 10 mM, too high for a good candidate. However, the inhibitory effect might be due to the DMSO rather than the presence of #3. When the concentration of DMSO reaches 6%, the activity is decreased by 2%. At the highest #3 concentration tested, 5% of DMSO was also present in the reaction so this might be the cause of inhibition. A co-crystallisation trail with #3 was set up and crystals obtained. They diffracted at 1.7 Å (at Diamond Light Source beamline I04) but the difference map did not show the presence of the compound. This would be consistent with the DMSO inhibitory effect.

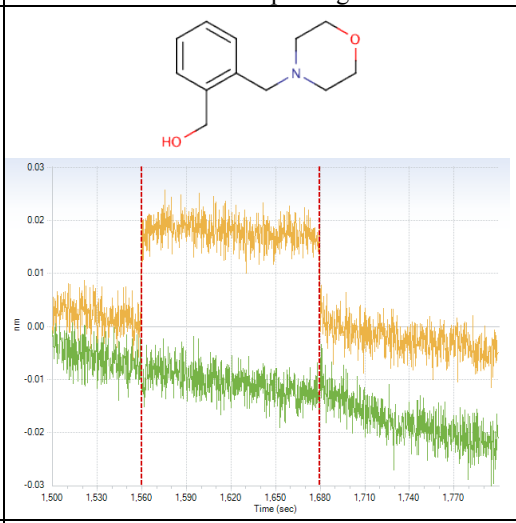
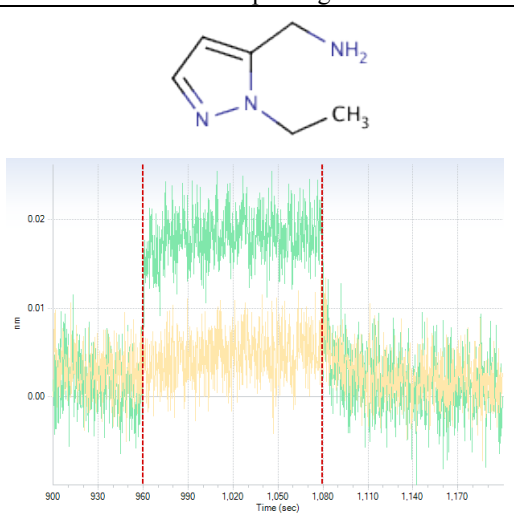
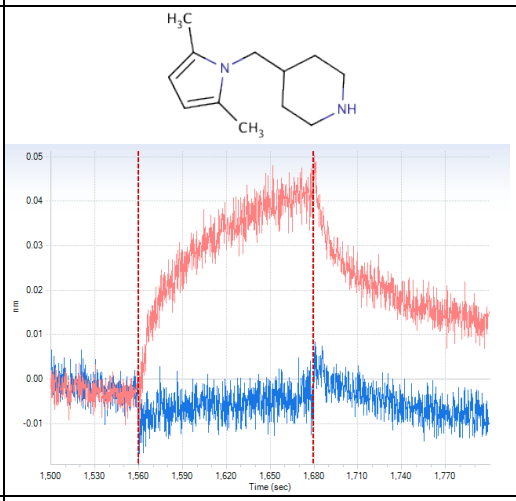
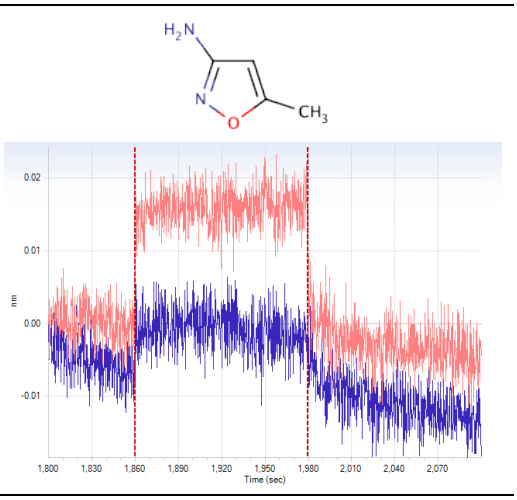
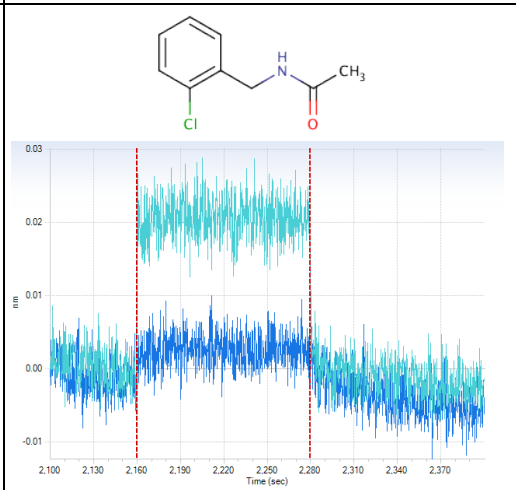
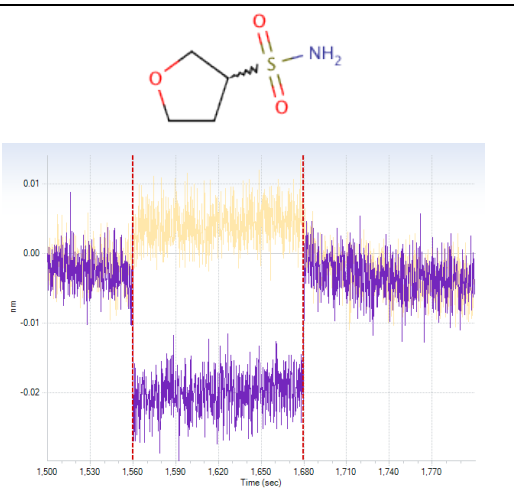
#	Structure / Spectrogram	#	Structure / Spectrogram
1		4	
2		5	
3		6	

Table 4.9: Fragment screening hit compounds and their binding signal (Δλ against time in seconds). Two binding curves are represented; one from TEV protease (showing no binding) and other showing *Bp*Ddl binding.

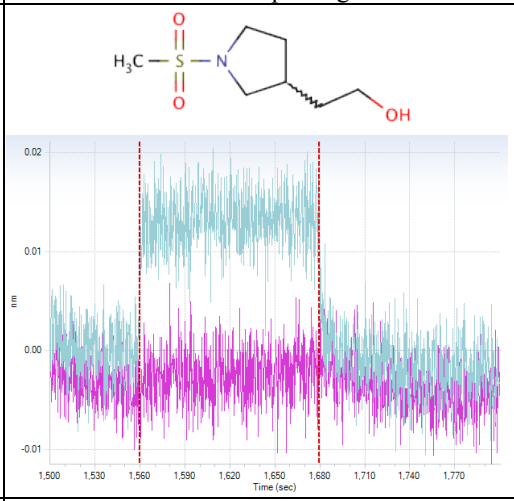
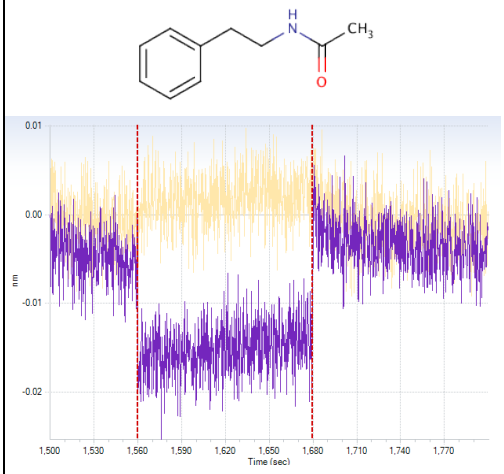
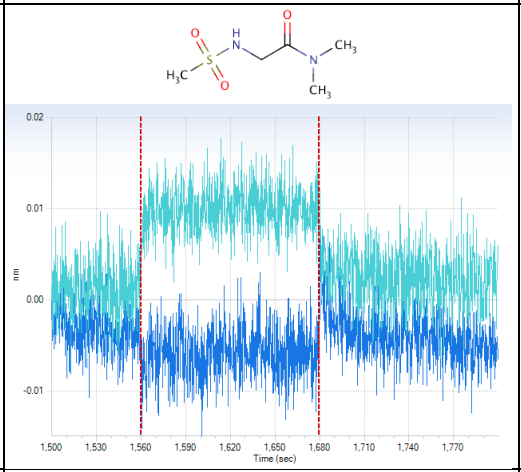
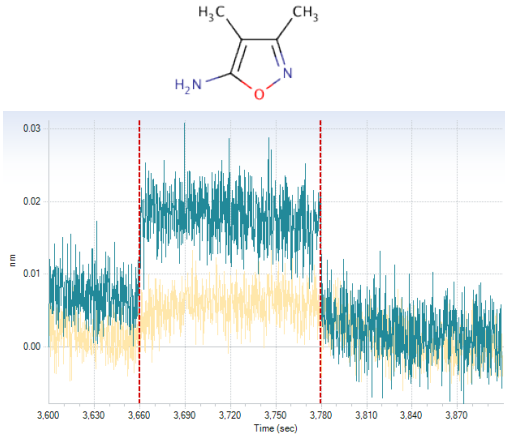
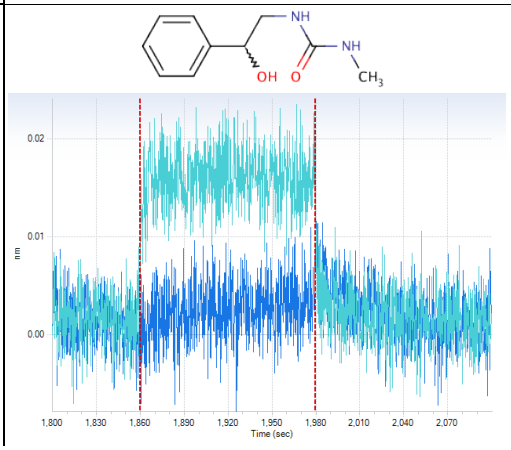
#	Structure / Spectrogram	#	Structure / Spectrogram
7		10	
8		11	
9			

Table 4.9: Continuation.

4.3.6 HTP screens

An HTP enzymatic assay measures activity which provides a direct tool to search for inhibitors targeting the active site. Therefore, it reduces the probability of finding false positive inhibitors.

The ELF proposed criteria that every protein should fulfil in order to assess compound library screening. The criteria includes: homogenous assay available for a 384-well plate format, in a maximum volume of 30 µl, maximum protein concentration of 100 µM, signal-to-noise ratio higher than three, a Z' score greater than 0.6, DMSO tolerance of 0.5%, read out stability for at least 1 hour and proteins not affected by freeze-thaw process. The HTP *Bp*Ddl assay satisfied these criteria.

4.3.6.1 EFL library screening

The HTP assay had to be optimised. Firstly, a test to check the protein activity and signal at different enzyme concentrations was performed (Figure 4.33). A concentration of about 0.2 nM was chosen for further experiments. In Figure 4.33, the data corresponding to 0.14 nM of *Bp*Ddl is the closest to the chosen concentration and it is linear for 5 minutes. To get a robust assay, the signal-to-noise ratio needs to be higher than 3. In the case of 0.023 nM of enzyme, this is achieved only after 20 min of reaction while for 0.138 nM, it is close to 3 after 5 minutes. Additionally, this assay was performed at saturating concentrations of D-Ala and ATP. The HTP screening was developed to target the substrate and co-factor pockets so D-Ala and ATP concentrations needed to be close to their K_m values. This implies the signal-to-noise ratio will be lowered.

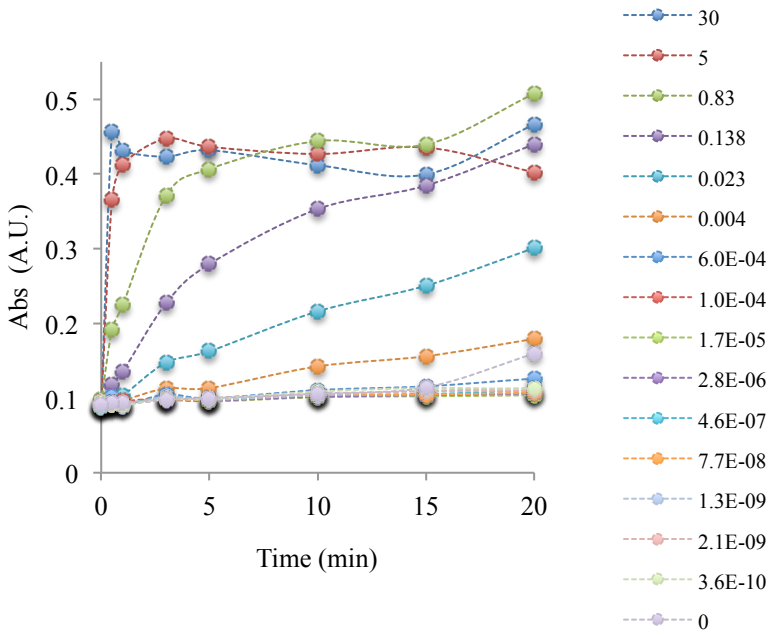


Figure 4.33: Enzyme titration experiment. The graph shows absorbance units (AU) against time (min). The different lines correspond to different enzyme concentrations (nM).

Once the enzyme concentration was chosen, the K_m experiments reported K_m values of 4 mM and 130 µM for D-Ala and ATP respectively (Table 4.8, section 4.3.4). V_{max} values have not been calculated as nmol of phosphate formed per minute because for every plate it was necessary to create a standard curve and there was no space in the plates for performing this experiment.

A DMSO tolerance experiment was performed (Figure 4.34) and the protein displayed a tolerance of 6% DMSO (only 2% inhibition of the enzyme activity). The final DMSO concentration during screening reached 0.075% so the assay has not been compromised.

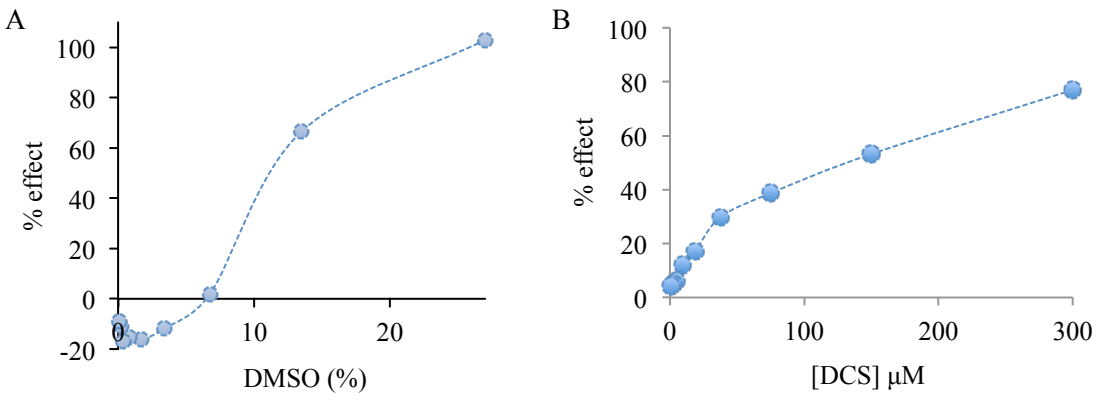


Figure 4.34: (A) DMSO dose-response curve to determine DMSO tolerance. (B) DCS dose-response curve.

A library of 6775 compounds was obtained from the ELF. This library is used by the ELF for HTP assay validation as it includes well known drugs, active compounds and aggregators. In each screening plate negative and positive controls are required. Ideally, the negative control should be the total inhibition of the enzyme by a known inhibitor. However, DCS did not show 100% inhibition (Figure 4.34). Instead, the negative control was performed without protein. Initially, the experiments were run in the absence of TCEP and Tween-20. To target the substrate and co-factor binding sites, a concentration of D-Ala and ATP close to the K_m were used. To confirm that these assay conditions were favourable, a time course experiment was performed to check for linearity. It was performed for 92 minutes and the reaction was still linear at 80 minutes. The Z' score at this time point was 0.8 and the ratio signal-to-noise 5. These matched the ELF criteria and the assay was ready for use. The results are summarised in Figure 4.35.

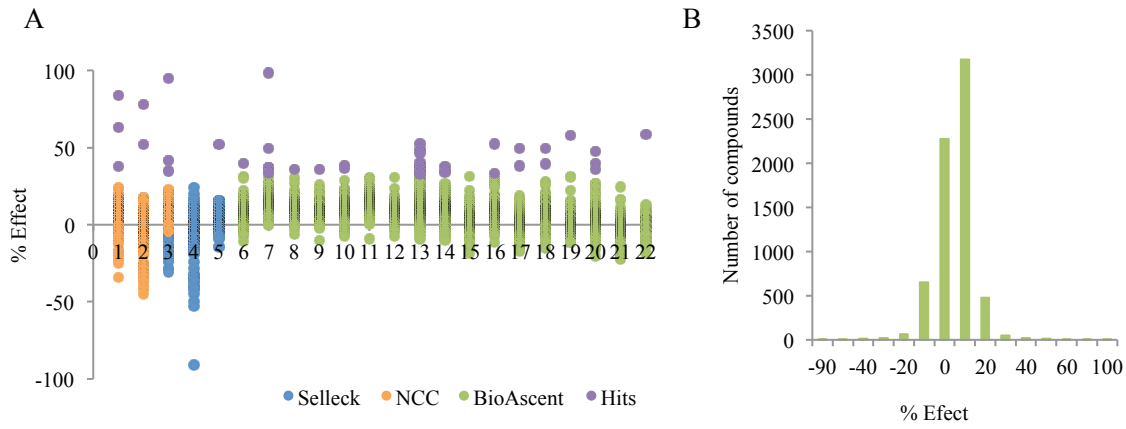


Figure 4.35: (A) Result of the ELF library screening represented as the percentage of inhibitory effect against the plate number. Compounds from different libraries are coloured in blue (Selleck), orange (NCC) and green (BioAscent), and hits are marked in purple. (B) Result of the ELF library as the compounds distribution along the percentage of effect.

After screening the ELF library, 74 probable hits were selected (Figure 4.35) and re-tested using the BIOMOL Green assay. To check that compounds were not interfering with the BIOMOL Green assay mixture two additional tests were done. The same 74

compounds were subject to the BIOMOL Green assay in the absence of enzyme or substrates, and separately, in the presence of BIOMOL Green and 70 μM of phosphate. For these two plates, the positive controls contained assay buffer and 70 μM of phosphate and negative control assay buffer without any supplemented phosphate. Results showed no interference by any of the compounds with the BIOMOL reaction. From the 74 hits, 61 were false positives leaving 13 compounds for further testing.

Compound ID	Compound concentration (μM)							IC_{50}
	30	10	3.33	1.11	0.37	0.12	0.04	
BCC0038183	24.10	-1.85						-
BCC0125015	30.23	2.12						-
BCC0112154	41.76	2.84						-
BCC0115825	48.60	30.59						-
BCC0013632	8.60	4.28						-
BCC0028871	42.12	8.24						-
BCC0032689	14.73	12.93						-
BCC0047923	44.64	35.27						-
S4059	82.48	41.40	9.68	4.28	1.76	2.84	8.24	12.88
SAM001247071	97.61	96.53	98.69	81.04	7.16	11.85	2.84	0.98
SAM001247083	85.00	58.33	26.98	5.72	8.24	-2.21	4.64	7.59
SAM002264651	36.35	16.53	-2.57	-2.21	-0.41	5.72	3.92	44.71
SAM002554903	87.16	75.27	15.81	-4.01	1.40	1.40	0.32	5.89

Table 4.10: Results from the dose-response experiment for 13 hit compounds. Compound IDs for NCC (BCC0XXXXXX), Selleck (SXXXX) and BioAscent (SAM00XXXXXXX) libraries.

The dose-response experiment was performed in the absence and presence of 1 mM TCEP and 0.01% (w/v) Tween-20. In the first case, IC_{50} values were calculated for five of the compounds. The other eight had a low stock volume and only two concentrations (20 and 30 μM) were tested. However, at the maximum concentration (30 μM) the percentage effect was less than 50% indicating they are not good candidates. Indeed, from the other five, four presented inhibition greater than 80%. The dose-response curves and the compound structures are in Figure 4.36. The IC_{50} and the results are shown in Table 4.10. When the same experiment was carried out in the presence of 1 mM TCEP and 0.01% (w/v) Tween-20, no inhibition was observed. This suggests the redox environment affects the stability of the compounds and/or they are protein

aggregators. Looking at the structure of the compounds in Figure 4.36, it is very likely they are aggregators (Baell and Holloway, 2010). The dose-response profile from SAM001247071 is also characteristic of an aggregator dose-response (Shoichet, 2006). Due to these results, it is concluded that no leads were obtained from this screen. Results provide confirmation of the reliability of the assay since aggregators were found in the screening.

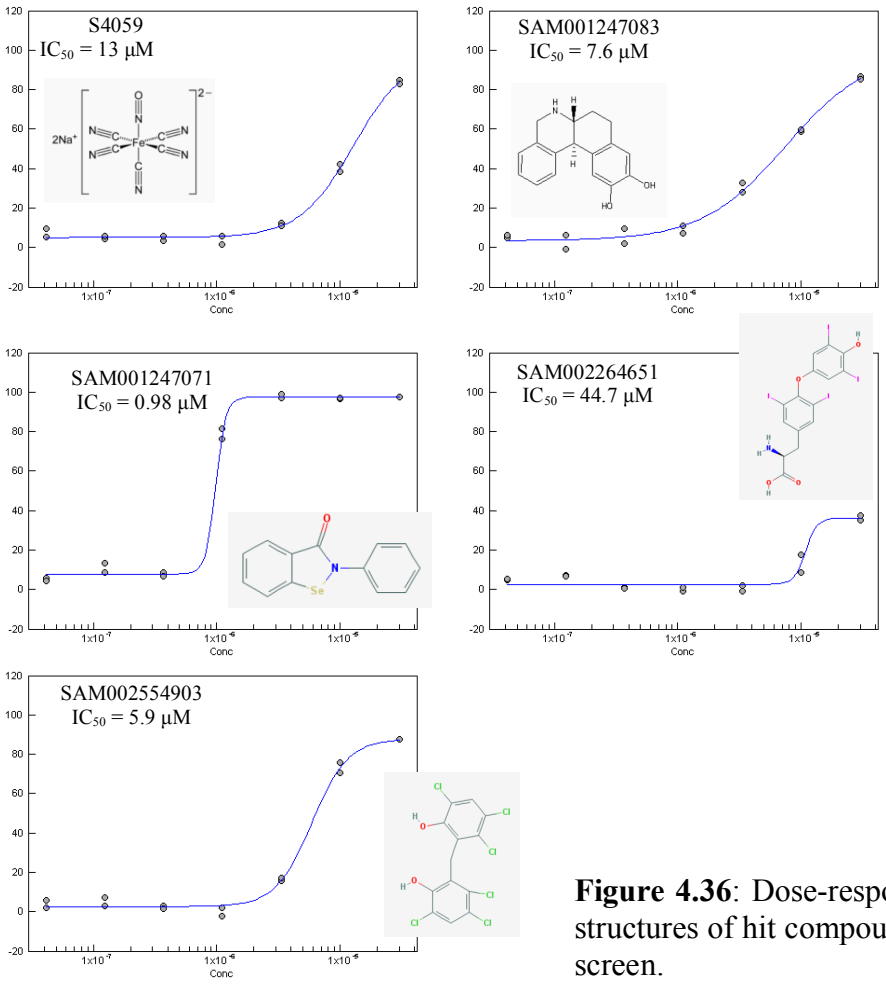


Figure 4.36: Dose-response curves and structures of hit compounds for the ELF screen.

4.3.6.2 DDU small diversity set screening

The optimisation carried out at the ELF provided a reliable HTP assay but no good inhibitors were found using their library. To check whether the D-Ala binding pocket is druggable, the DDU diversity set screen was selected. Prior to the realisation of this screening, a peer review of the HTP assay was performed with the screening team at the

DDU. Some changes were made and a new optimisation of the HTP assay was carried out to set up new screening conditions.

The starting point was the performance of the assay at higher enzyme concentration and a pH closer to the physiological condition (pH 7.4). At this pH, using 0.2 nM enzyme, K_m calculations for ATP and D-Ala were carried out (Table 4.8). In the case of the diversity set, only the D-Ala pocket was prioritised as the library target. The initial assay conditions included K_m concentration for D-Ala and saturating concentration of ATP. After performing the time course experiment at different enzyme concentrations (1, 2 and 3 nM), only the reaction with 1nM was linear for 10 minutes. In order to get a longer assay time scale, the concentration of D-Ala was decreased down to 1 mM. In this case, the reaction was linear for 20 minutes, the signal-to-noise ratio 6 and the Z' score 0.8-0.9.

The screen was then performed with 1 mM D-Ala, 500 μM ATP, and incubated for 12 minutes. To minimize false positives (avoid aggregators) the assay was performed in the presence of TCEP and Tween-20. The screening results are represented in Figure 4.34. From the 15667 compounds tested four hits were selected (showing 30% inhibitory effect, Figure 4.37). A dose-response experiment was carried out for these compounds but none of them showed any inhibitory activity.

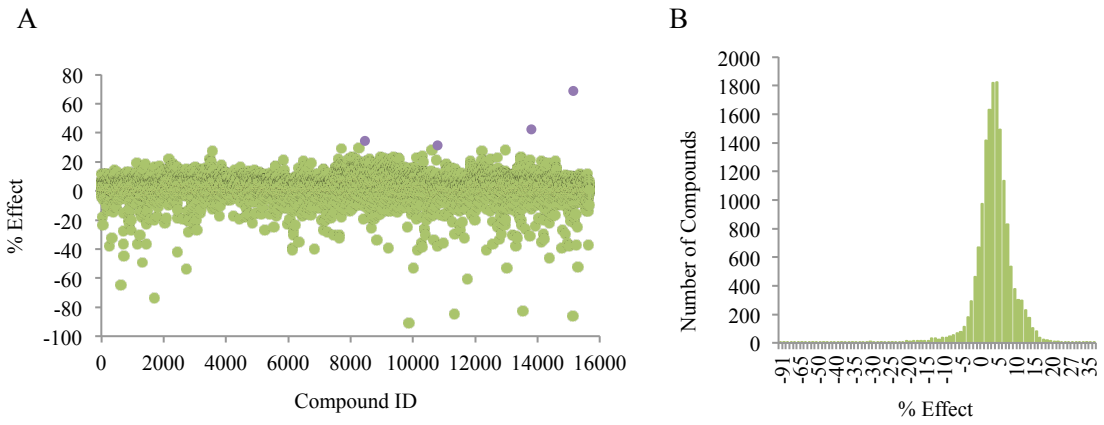


Figure 4.37: (A) Result of the DDU diversity set represented as the percentage of inhibitory effect against the compound ID. Hits are coloured in purple. (B) Result of the DDU diversity as the compounds distribution along the percentage of effect.

4.4 Conclusions

Having selected Ddl as a potential drug target in *B. pseudomallei* and *F. tularensis*, and with the results of the Dstl labs showing essentiality for *B. pseudomallei* growth, we prioritised this enzyme for assessment for drug discovery. Firstly, protein production and purification of *Bp*Ddl and *Ft*Ddl were developed and optimised. Well diffracting *Bp*Ddl crystals were obtained in the presence of various ligands and structures determined at 1.3 and 1.5 Å resolution in the presence of AMP and AMP+D-Ala—D-Ala respectively. These structures were used as a template for druggability assessment and drug design. Additionally, the analysis of different *Bp*Ddl structures suggested the protein was being purified in the presence of AMP. This was confirmed using an AMP detection assay. This means kinetic parameter determination, screens and crystallisation were performed on the AMP-bound conformation of the protein instead of the apo-form. This had not been reported previously and therefore we questioned whether this might also apply in other Ddl studies. So far, no information about co-purification with AMP has been noted. The *Bp*Ddl activity was measured using two different enzymatic

assays and showed similar kinetic parameters to other Ddl proteins. Considering the sample has already AMP bound, it is necessary to note that a retarded phase during the assay performance was observed, probably due to the presence of the ligand, and the $K_{m,ATP}$ examination and comparison with other enzymes needs to take into account these observations.

HTP binding and enzymatic assays were developed to conduct DDU and ELF compound library screens. Despite screening more than 22500 compounds, no good inhibitors were identified. A team at AstraZeneca reported a compound library screen for *Streptococcus mutants* Ddl but no lead compounds were found (Tommasi *et al.,* 2015), which is consistent with our results.

The overall assessment is that *Bp*Ddl is a challenging target and a HTP screen approach might not be the best strategy to identify new inhibitors for this protein. New strategies need to be developed for drug discovery for this target. A rational structure-based design of compounds targeting *Bp*Ddl might be carried out. This could include compounds targeting the D-Ala pocket but with an extension to the ATP pocket, or molecules disrupting or preventing the ω-loop conformational change.

PART FIVE

CASEINOLYTIC PROTEASE SUBUNIT P

5.1 Introduction

The caseinolytic proteases (Clp) are present in all bacteria with orthologues in eukaryotes (Brötz-Oesterhelt and Sass, 2014). Their role is in maintaining protein homeostasis in the cell, especially under stress conditions. They can indirectly modulate gene expression, cell motility and cell division, and they are involved in pathogenesis by promoting virulence factor expression, removing stress-damaged proteins and contribute an important role in the production of exotoxins (Brötz-Oesterhelt and Sass, 2014; Frees *et al.,* 2014)

The Clp protease complex displays a cylindrical shape with an internal chamber where the proteolysis of unfolded proteins is carried out (Brötz-Oesterhelt and Sass, 2014). Different Clp protease complexes are constructed of a core, formed by the protease ClpP, surrounded by accessory proteins (AAA+ chaperones ClpA, ClpX and ClpC) that regulate the specificity of the proteolysis (Alexopoulos *et al.,* 2012). ClpP is an unspecific serine protease containing the catalytic triad Ser-His-Asp and forms seven-subunit rings that associate to generate a tetradecamer structure. Each ClpP monomer displays proteolytic activity, which depends on the formation of the tetradecamer complex (Brötz-Oesterhelt and Sass, 2014). The complex presents two axial pores for the entrance of peptide substrates to the active site (Geiger *et al.,* 2011). In the absence of the regulatory proteins ClpA/X/C, the axial pore size allows small unfolded peptides to access the catalytic chamber. To speed up the reaction and increase specificity, the regulatory proteins bind either directly to the targeted protein or to other chaperones

carrying the target. They unfold the proteins and induce the aperture of the proteolytic core (Brötz-Oesterhelt and Sass, 2014). Once the reaction has been carried out, the conformation of the ClpP core changes and generates pores in the equatorial plane to release the cleaved peptides (Figure 5.1, Geiger *et al.,* 2011; Lee *et al.,* 2011). The ClpP structure is highly conserved among different organisms, constituting the ClpP folding family (Yu and Houry, 2007; see section 5.3.3 for further details).

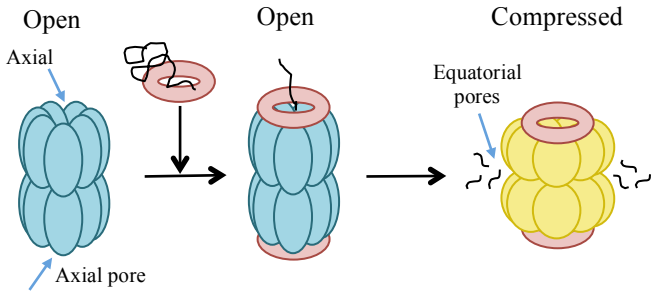


Figure 5.1: Scheme of Clp protease function. The adaptors/chaperones ClpA/X/C (red) bind to the misfolded protein (black) for degradation. They interact with the ClpP complex (blue) and unfold the protein so it can be proteolysed. The conformation changes (yellow) and products are released from pores in the equatorial plane.

Clp proteases have been studied as antibacterial drug targets and different compounds that alter ClpP function have been found (Brötz-Oesterhelt and Sass, 2014). Depending on the effect of the compounds they can be used as antivirulence or antibacterial agents. Inhibitors, such as lactone D3 and U1, prevent virulence of *Staphylococcus aureus* (Böttcher and Sieber, 2008) and *Listeria monocytogenes* (Böttcher and Sieber, 2009). On the other hand, activation of the complex leads to unspecific proteolysis and therefore, activators act as antibiotics. This is the case for acyldepsipeptides (ADEP) (Sowole *et al.,* 2013).

A total of 31 ClpP structures have been determined including apo and ligand-bound with resolutions from 1.8 to 2.4 Å. Two structures in the presence of the activator ADEP, from *Bacillus subtilis* (*Bs*ClpP, Lee *et al.,* 2010) and *E. coli* ClpP (*Ec*ClpP, Li *et al.,* 2010), one with the inhibitor diisopropyl phosphonate (*Bs*ClpP, Lee *et al.,* 2011), and three bound to small peptides, from *Ec*ClpP (Szyk and Maurizi, 2006) and *H. pylori*

(*Hp*ClpP, Kim and Kim, 2008) have been characterised. The two different conformations for ClpP, open and compressed, have been structurally characterised for *Bs*ClpP (Lee *et al.,* 2011), *M. tuberculosis* (*Mt*ClpP, Ingvarsson *et al.,* 2007) and *S. aureus* (*Sa*ClpP, Geiger *et al.,* 2011; Zhang *et al.,* 2011). An apo structure of *F. tularensis* ClpP (*Ft*ClpP, open conformation, 3P2L, 2.3 Å resolution) has been determined.

ClpP was selected as a potential target for *F. tularensis* and validation studies initiated. Two structures presenting the open and compressed conformations have been determined.

5.2 Experimental procedures

5.2.1 Recombinant protein production

The codon optimised gene encoding *Ft*ClpP (UniProt: Q5NH47) was purchased from GeneWiz and cloned into the pET15bTEV vector. Recombinant expression was carried out in *E. coli* BL21(DE3), using 1 mM IPTG for induction of expression and incubating the culture overnight at 20ºC. The first Ni-NTA column for the purification of *Ft*ClpP was performed using buffer A (20 mM Tris-HCl pH 7.4 and 200 mM NaCl) and buffer B (buffer A with 0.8 M imidazole). The His_6-tag was cleaved using TEV protease (1 mg per 10 mg of *Ft*ClpP) by incubating the samples at room temperature for 2h or at 4ºC overnight. Then a second Ni-NTA chromatography step was performed after dialysis into buffer A to separate the His_6-tagged and non-tagged proteins. To determine the quaternary structure, SEC using a HiLoad 16/60 superdex 200 prep grade column (GE Healthcare) and native-PAGE (Invitrogen) were performed in buffer A, with and without 10% glycerol. Calibration of the SEC column was carried out as in Part Three

using buffer A with and without 10% glycerol. Protein purity was assessed using SDS-PAGE (with Coomassie Blue stain) and MALDI-TOF-MS analysis. Protein concentration was determined using the predicted ε(*Ft*ClpP) = 8940 $M^{-1}cm^{-1}$ (value obtained from PROTPARAM, Gasteiger *et al.,* 2003) and the Lambert-Beer equation.

5.2.2 Crystallographic analysis

5.2.2.1 Crystallisation

*Ft*ClpP samples were concentrated to 20 mg ml^{-1} and commercially available crystallisation conditions were tested. These included The Classics Suite from Qiagen, JSCG plus and PEGs screen from Molecular Dimensions. Conditions were assessed using 1:1 and 1:2 protein solution:reservoir ratios (final drop volumes of 0.2 and 0.3 µl respectively) in sitting drop 96-well plates. Diffracting crystals were obtained from The Classics Suite A4 (5% (v/v) isopropanol and 2 M $(NH_4)_2SO_4$) and B4 [0.2 M NaCl, 30% (v/v) MPD (2-methyl-2,4-pentanediol) and 0.1 M sodium acetate pH 4.6] conditions. Optimisation was performed using 1:1 protein solution:reservoir ratio in a final drop volume of 4 µl and using hanging drop plates.

5.2.2.2 Diffraction measurements and structure determination

Crystals from the B4 condition diffracted to 2.8 Å resolution. Data were collected using the in-house X-ray generator (Rigaku M007HF X-ray generator with a Saturn 944HG+ CCD detector), a 10º θ offset, 10 seconds of exposure time and an oscillation range of 0.5º (covering 115º collected). Data integration and analysis were performed using iMOSFLM (Leslie and Powell, 2007) and AIMLESS (Evans, 2011). The crystal has a space group $P2_12_12$ and unit cell dimensions of a=113.5, b=125.9, c=96.95 Å. The Matthews coefficient calculated with the CCP4i suite (Winn *et al.,* 2011) indicated there were seven monomers in the asymmetric (V_m value of 2.2 $Å^3$ Da^{-1} and bulk solvent of

45%). The *Ft*ClpP structure 3P2L (2.3 Å resolution, space group $P2_12_12$ and unit cell dimensions a=120.52, b=128.82, c=98.03 Å), also containing seven subunits in the asymmetric unit, was used as the search model for molecular replacement. Firstly, a monomer from 3P2L was employed and a partial solution was obtained using PhaserMR (McCoy *et al.,* 2007). Only three subunits were found and clashes were noticed. Molecular replacement was then carried out with the heptamer but the high number of clashes for the probable solutions did not support the packing and no acceptable solutions were obtained. Analysing the electron density maps from the initial PhaserMR results, suggested a change in the conformation. Accordingly, the structure was solved using the subunit A from 3P2L without residues 128 to 149 as the model, and it was called *Ft*ClpP:Form-I.

Diffraction data were also collected at ESRF beamline BM30-A. Crystals grown from an optimised The Classics Suite B4 condition, diffracted to 1.5-1.8 Å. Four crystals were used for data collection and then analysed using XDS (Kabsch, 2010) and AIMLESS. All crystals had the same space group ($P2_12_12$) and two crystal forms can be identified with cell dimensions of a=98.49, b=128.23, c=355.18 or a=117.90, b=128.20, c=98.26 Å. Two data sets, containing fewer reflection overlaps, were chosen for structure characterisation and named *Ft*ClpP:Form-II and *Ft*ClpP:Form-III. Matthews coefficients were determined and suggested the presence of 7 (V_m of 2.41 $Å^3$ Da^{-1} and bulk solvent 50%) and 21 (V_m of 2.43 $Å^3$ Da^{-1} and bulk solvent 50%) subunits in the asymmetric unit in *Ft*ClpP:Form-II and *Ft*ClpP:Form-III, respectively. *Ft*ClpP:Form-II was solved using PhaserMR with subunit A from the 3P2L model. In the case of *Ft*ClpP:Form-III, monomer, heptamer and tetradecamer models of the refined *Ft*ClpP:Form-I and *Ft*ClpP:Form-II were tried as search models but no good solutions were found. This might indicate problems with the data or the presence of a different

conformation and further analysis is required. The general refinement procedure for *Ft*ClpP:Form-I and *Ft*ClpP:Form-II data sets started with a first round of rigid body refinement using REFMAC (Murshudov *et al.,* 2011) followed by cycles of restrained refinement. Electron density and difference maps were inspected, and modifications to the model including addition of rotamers, waters and ligands were made using COOT (Emsley *et al.,* 2010). Tight non-crystallographic symmetry restraints were used at the beginning of the refinement and removed gradually. When inspection of the maps suggested no more changes were justified, and no significant changes in the R_{free} and R_{work} were observed, the refinements were ended. The electron density in some areas of the models did not support a reliable addition of residues. Missing sections of the peptide chains are detailed in Table 5.1.

***Ft*ClpP:Form-I**							
Subunit	A	B	C	D	E	F	G
N-terminus	1-4	1-5, 13-17	1-5, 12-15	1-4, 12-20	1-5, 12-20	1-4, 13-19	1-4, 13-18
Handle region	132-139	131-141	133-141	133-141	132-142	131-141	133-141
C-terminus	196-201	196-201	196-201	196-201	197-201	194-201	197-201

***Ft*ClpP:Form-II**							
Subunit	A	B	C	D	E	F	G
N-terminus	1-5	1-4, 11-19	1-3, 13-20	1-6, 11-20	1-5*, 11-19	1-4	1-4
Handle region	-	-	-	-	-	-	-
C-terminus	199-201	200-201	-	299-201	-	189-201	199-201

Table 5.1: Missing residues from models *Ft*ClpP:Form-I and *Ft*ClpP:Form-II. (*) In residue number 5, only density supporting the presence of the main chain is visible and Ala was placed instead of Asn.

5.2.2.3 Further structural analysis

The models were analysed using MolProbity (Chen *et al.,* 2010). The secondary structure was inspected with DSSP (Touw *et al.,* 2015) and the search for structural homologues and comparisons were performed with the DALI server (Holm and Rosenstrom, 2010). Images were prepared using PyMol (www.pymol.org) and ALINE (Bond and Schüttelkopf, 2009). The interface surface area was calculated with PISA

(Krissinel and Henrick, 2007). Amino acid sequences were analysed with Clustal Omega (Sievers *et al.,* 2011) and Jalview (Waterhouse *et al.,* 2009), and structural alignments were performed using PyMol.

5.3 Results and discussion

5.3.1 Recombinant protein production

The expression vector was sequenced and no mutations were found. Protein production was achieved and yielded 10 mg of *Ft*ClpP per liter of culture. *Ft*ClpP does not contain any tryptophan and therefore, the gels were stained with Coomassie Blue and showed *Ft*ClpP presented a high level of purity. The predicted molecular weight of the *Ft*ClpP subunit is 22.1 kDa and formation of heptamers or tetradecamers has been reported (Lee *et al.,* 2010). During the characterisation of *B. subtilis* ClpP, formation of the tetradecamer was observed with SEC in the presence of 10% glycerol. In the absence of

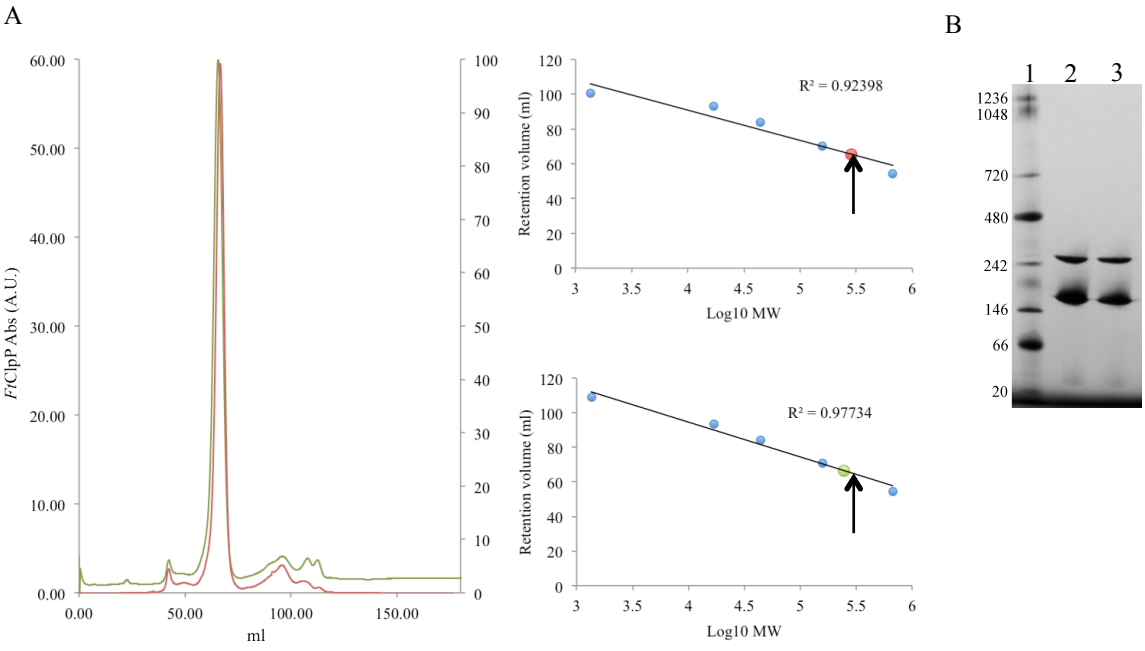


Figure 5.2: (A) SEC results in the absence (red) and presence (green) of 10% glycerol. The equilibration results, without (top) and with (bottom) glycerol, are shown in the graphs. Arrows indicate the expected value for a tetradecamer. (B) Native-PAGE 3-12%. Molecular weight standards are in lane 1. *Ft*ClpP was loaded in lanes 2 and 3 (7 and 5 ng of protein respectively).

glycerol, a peak corresponding to the molecular weight of a heptamer was observed (Lee *et al.,* 2010). The *Ft*ClpP quaternary structure determination was first investigated using SEC in the absence and presence of 10% glycerol (Figure 5.2A). In the first case, a peak corresponding with 245 kDa was obtained and used for crystallisation experiments. When glycerol was present, a peak of 286 kDa was identified. In both cases, these values suggested *Ft*ClpP is present as a tetradecamer in solution. Results from the native-PAGE (Figure 5.2B) showed bands of 292 kDa and 170 kDa which correspond with the formation of tetradecamers and heptamers. The differences in the molecular weight shown by SEC and native-PAGE can be due to limitations on the techniques, the complex shape and size, precision of the native-PAGE and errors in the SEC column calibration.

E. coli ClpP (*Ec*ClpP) is synthesized as a propeptide and cleavage of the first 14 residues has been described (Maurizi *et al.,* 1990). MALDI-TOF MS analysis was used to check whether *Ft*ClpP was also present as a propeptide. The result showed a molecular weight of 21.6 kDa (only 0.5 kDa lower than expected) and no degradation products are noticed (Figure 5.3). Also, a high level of purity of the purified *Ft*ClpP can be observed.

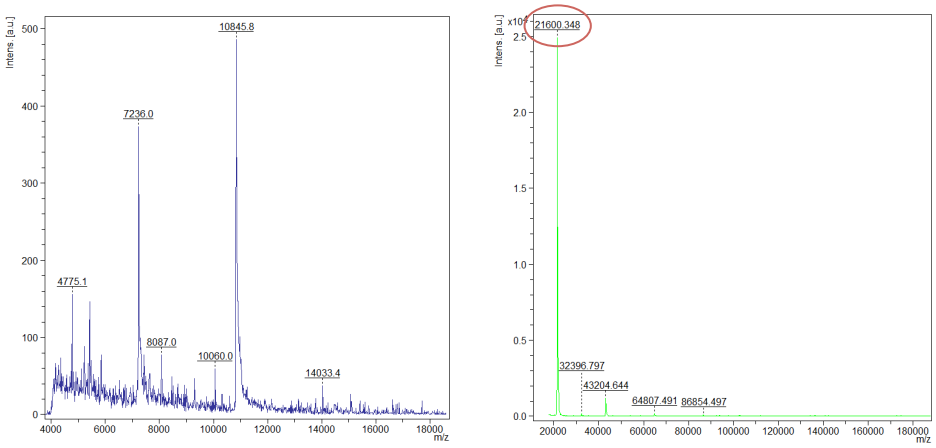


Figure 5.3: MALDI-TOF mass spectrometry result for *Ft*ClpP. The plots show the intensity of the signal (a.u.) against the mass:ion charge relation ship (m/z). On the left a run comprising molecular weights 4-20 kDa and on the right the results from 20-180 kDa run. Note the differences in the scales. The red circle indicates *Ft*ClpP peak.

5.3.2 Crystallisation and diffraction

Diffracting crystals were obtained from The Classics Suit condition B4 (Figure 5.4). The first structure (*Ft*ClpP:Form-I) was obtained from one of these crystals. Data were collected in-house and the resolution was 2.8 Å. Further optimisation gave bigger crystals with improved resolution. Bigger crystals (average dimensions 600 x 200 x 400 µm) grew from various optimisation conditions and two data sets were selected for structure determination and analysis (*Ft*ClpP:Form-II and *Ft*ClpP:Form-III). ClpP has been shown to have two conformational states; the open and compressed conformations. *Ft*ClpP:Form-I structure shows a compressed conformation while *Ft*ClpP:Form-II is the open form (like the molecular replacement model 3P2L). No information about the *Ft*ClpP:Form-III conformation is presented until the structure is solved. The crystallographic data for the three crystals are shown in Table 5.2.

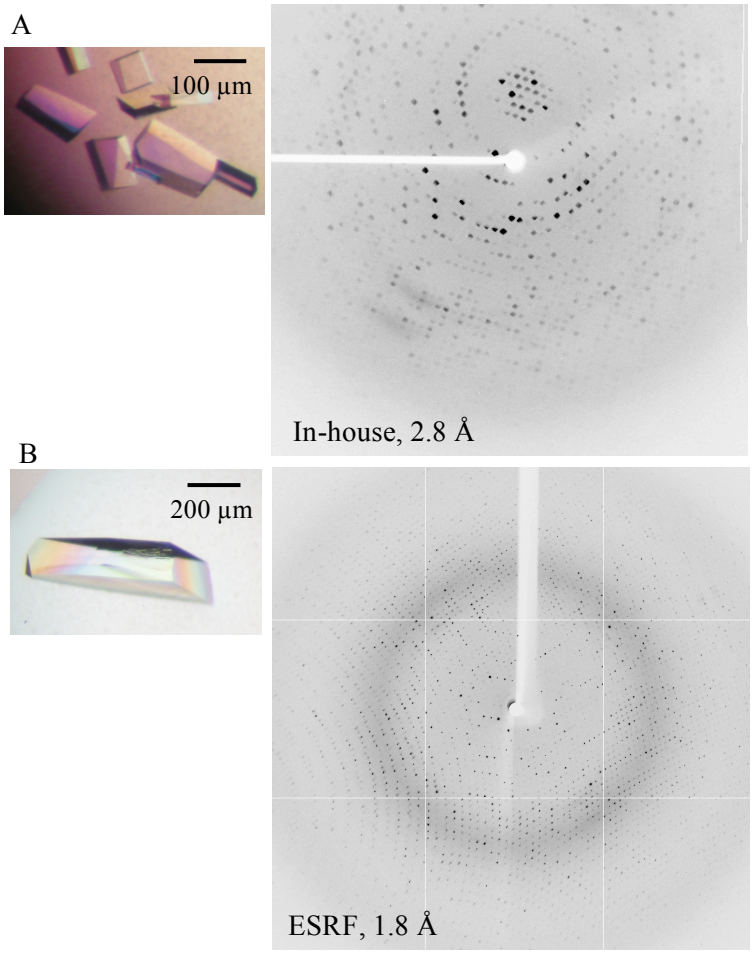


Figure 5.4: *Ft*ClpP crystals and diffraction patterns (A) The Classics Suite condition B4. (B) Result after the optimisation of condition (A).

Structure	***Ft*ClpP:Form-I**	***Ft*ClpP: Form-II**	***Ft*ClpP: Form-III**
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Wavelength (Å)	1.5418	0.9799	0.9799
Unit cell dimensions *a*, *b*, *c* (Å)	113.5 125.9 96.9	117.9 128.2 98.3	98.5 128.2 355.2
Resolution range[a] (Å)	63.61 - 2.84	47.04 - 1.90	47.60 - 1.8
No. reflections	90264	838386	2376413
Unique reflections	32741	117661	405163
Completeness (%)	97.8 (98.8)	100 (100)	97.9 (84.0)
R_{merge}[b]	0.098 (0.344)	0.085 (0.460)	0.061 (0.293)
Redundancy	2.8 (2.0)	7.1 (7.1)	5.9 (3.3)
< I/σ(I)>	5.9 (2.2)	18.5 (4.4)	16.3 (3.3)
Wilson *B* (Å^2)	15.6	10.8	10.1
Mosaic spread (º)	0.9	0.2	0.2
R_{work}[c]/R_{free}[d]	0.2437 / 0.3281	0.2008 / 0.2694	-
DPI[e] (Å)	0.4936	0.1653	-
R.m.s.d. bond lengths (Å) / angles[f] (º)	0.006 / 1.150	0.018 / 1.875	-
Average *B*-factors (Å^2)	36.27	20.94	-
Protein residues	1329	2436	-
Water molecules	-	1085	-
Ligands	-	5 MPD, 7 acetate	-
Ramachandran analyses			-
Favoured regions (%)	90.4	96.9	-
Allowed regions (%)	98.7	99.7	-

Table 5.2: Crystallographic statistics. [a.] Values in parentheses refer to the highest resolution shell. [b.] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. [c.] $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. [d.] R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations. [e.] Diffraction Precision Index (Cruickshank, 1999). [f.](Engh and Huber, 1991).

5.3.3 ClpP structure

Analysis of the *Ft*ClpP structure has been carried out using *Ft*ClpP:Form-I and *Ft*ClpP:Form-II which both contain seven subunits (A, B, C, D, E, F and G) in the asymmetric unit.

Each monomer presents the characteristic ClpP family fold consisting of eleven β-strands and seven α-helices arranged in head and handle regions (Figure 5.5). The handle is formed by α5 and β9 and the active site, Ser101-His126-Asp175, is located at the edge of the head and handle regions (Figure 5.5B). The β-strands 1 and 2 are not ordered in subunit C (*Ft*ClpP:Form-I) and G (*Ft*ClpP:Form-II). A search of the bacterial ClpP sequences (excluding *Francisella sp.*) in UniProt revealed 405 reviewed/curated entries. The amino acid alignment showed 40-70% identity with *Ft*ClpP and conservation of the three active site residues in all sequences. The sequence regions corresponding with the secondary structure elements are also highly conserved with the exception of the C-terminal region of α5 (not shown). Figure 5.6 shows the alignment of *Ft*ClpP with three orthologues (*Bs*ClpP, *Ec*ClpP and *M. tuberculosis* ClpP). The sequence identity between *Ft*ClpP and *Bs*ClpP is 45% and about 60% in the case of *Ec*ClpP and *Mt*ClpP. Comparing the sequences with the human mitochondrial ClpP, residues are highly conserved sharing 38% identity with the *Ft*ClpP. This could present issues in developing a specific antibacterial drug.

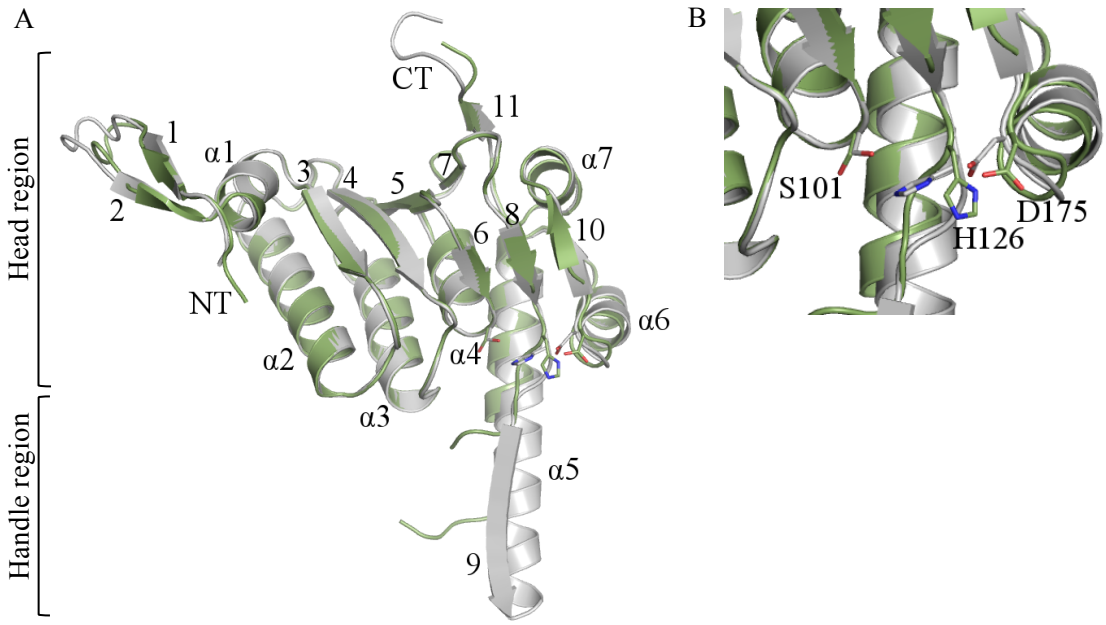


Figure 5.5: (A) Cartoon representation of the C_α alignment of monomers from *Ft*ClpP:open (grey) and *Ft*ClpP:compressed (green). α-helixes and β-strands are named like in Figure 5.6. Catalytic triad is represented in sticks. (B) Active site pocket showing the catalytic triad residues.

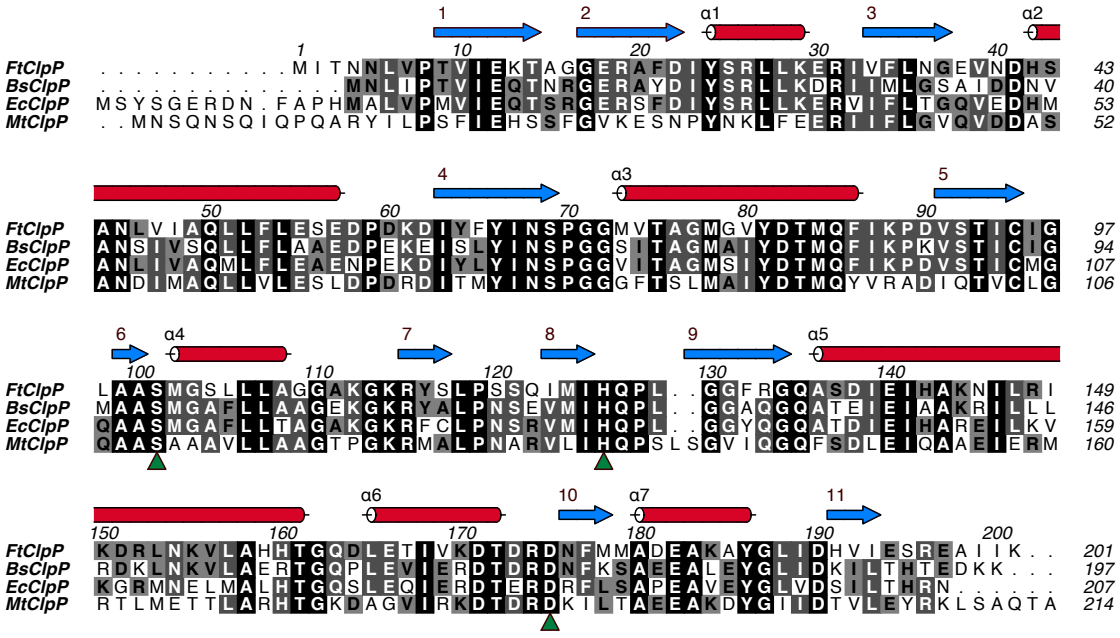


Figure 5.6: Sequence alignment of *Ft*ClpP (UniProt code Q5NH47), *Bs*ClpP (P80244), *Ec*ClpP (P0A697) and *Mt*ClpP (P9WPC3). Secondary structure elements are represented as red cylinders (α-helices) and blue arrows (β-strands). Active site residues are marked with green triangles.

5.3.3.1 ClpP complex analysis

The physiological arrangement of ClpP is a complex of 14 subunits by the association of two heptamer rings. *Ft*ClpP structures have the heptamer in the asymmetric unit and the tetradecamer is formed by a crystallographic two-fold rotation axis. The heptamer complex is formed by interactions of β-strands (3 and 4) and two loops (between β5 and β6, and β7 and β8) with α-helixes (α2, α3, α4 and α5) of the adjacent subunit. The tetradecamer formation occurs by the antiparallel hydrophobic interactions between the two α5, and hydrogen bonds between main chains of both β9 (Figure 5.7). The 14 active sites are located in the interface between the handle and the head region and the pockets form two grooves in the equatorial plate of the complex (Figure 5.7B). Two conformations have been described for ClpP; open (*Ft*ClpP:Form-II) and compressed (*Ft*ClpP:Form-I). The conformational change from open to compress allows the substrates to be released and is due to the disorganisation of the handle. This allows the head regions from one heptamer ring to approach the other, generating a more compressed conformation. The structural rearrangement produces changes at the active site pocket (Figure 5.5B) and the loss of the active site groove organisation (Figure 5.7B). All this ultimately leads to the formation of equatorial pores where release of the products occurs (Lee *et al.,* 2011). Analysis of the formation of these quaternary structures has been made using PISA. In the case of *Ft*ClpP:Form-I, 35 ± 2% of the monomer accessible surface area is involved in the tetradecamer formation and 43 ± 1% in *Ft*ClpP:Form-II. In the complex, the buried surface areas are about 47710 $Å^2$ and 46390 $Å^2$, respectively which corresponds to a stable assembly (see Part Four).

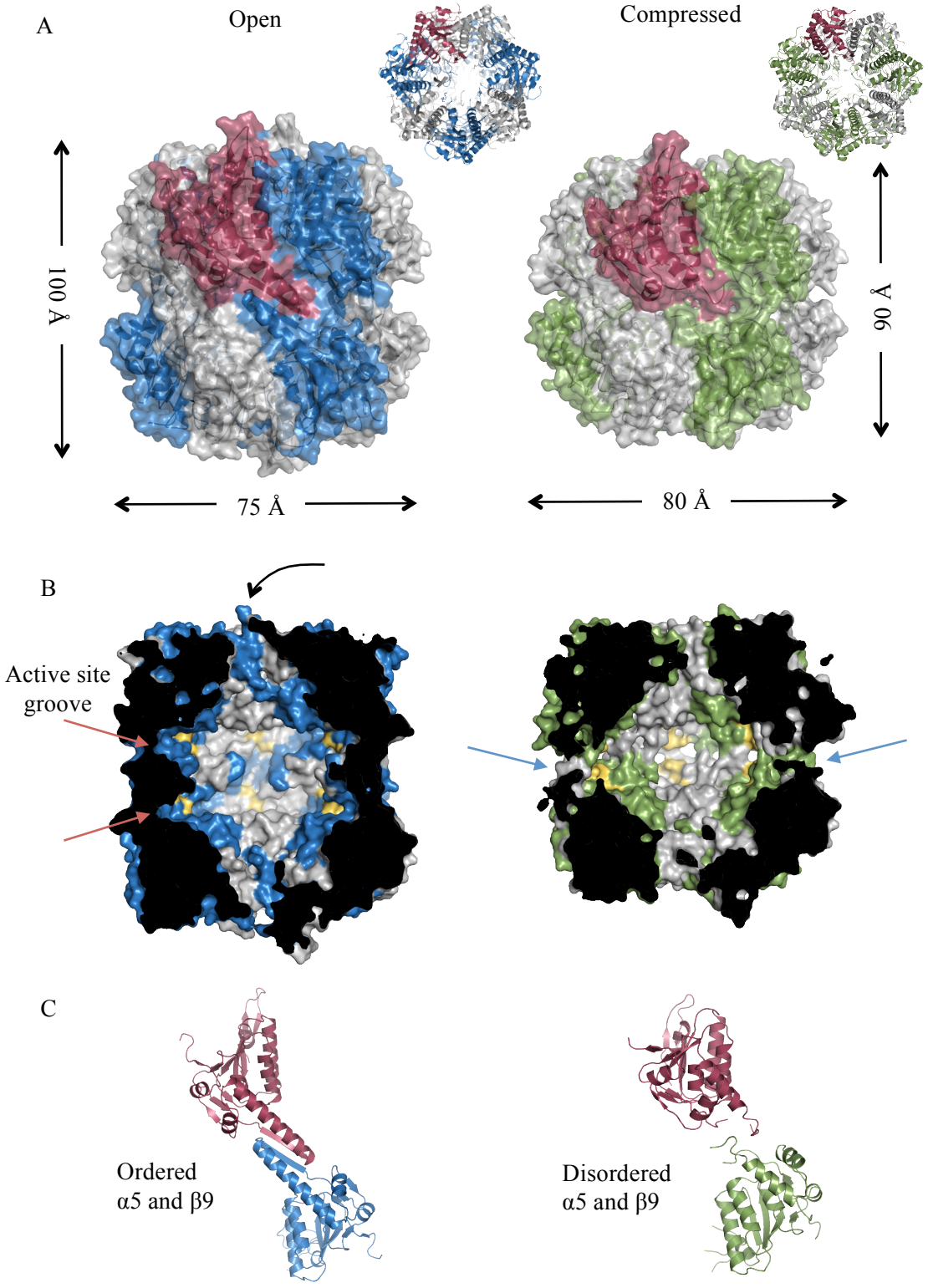


Figure 5.7: Structural insights into *Ft*ClpP:Form-II (grey and blue), and *Ft*ClpP:Form-I (grey and green). The seventh subunit of the heptamer is coloured in red. (A) Ribbon representation of the complex top view, and surface overlaid with ribbon representation of the lateral view. The arrows and numbers indicate the dimensions of both conformations. (B) Surface representation of the catalytic chamber. The black arrow indicates one of the translocation sites for the substrates. The red arrows mark the active site grooves and the blue arrows the pores for the products release. The active site amino acids are coloured in yellow. (C) Ribbon representation of the interaction between subunits from different rings that forms the tetradecamer.

Structural C_α overlay of *Ft*ClpP:Form-I subunits indicates high similarity among subunits (r.m.s.d. average of 0.38 ± 0.04 Å). The same observation applies to *Ft*ClpP:Form-II subunits with an r.m.s.d. average of 0.21 ± 0.03 Å. Performing pairwise overlaps between *Ft*ClpP:Form-I and *Ft*ClpP:Form-II subunits (0.38 ± 0.04 Å), minor differences reside in β1, β2 and the C-terminus positions. Even when the heptamers are compared, there are no major structural differences (r.m.s.d. of 0.78 Å). However, the overlay of the tetradecamer complexes shows an r.m.s.d. value of 4.29 Å what is consistent with the large conformational change observed. A search for structural homologues among published structures showed a high homology with all the ClpP proteins structurally characterised (30-70% identity, r.m.s.d. 0.5-2.5 Å), and shares a similar monomer conformation with enoyl-CoA-hydratase from *Mycobacterium smegmatis* (18% identity, r.m.s.d. 2.1 Å, Baugh *et al.,* 2015), and naphthoate synthase from *Synechocystis sp* (12% identity, 2.6 Å, Sun *et al.,* 2012).

5.4 Conclusions

ClpP was considered as a potential drug target for *F. tularensis*. In contrast to other systems, the search of an activator to generate unspecific proteolysis, rather than an inhibitor, was the chosen strategy. Production of highly pure *Ft*ClpP samples has been achieved and crystallisation optimised. Structures of *Ft*ClpP in the open and compressed stage were obtained and indicated the conformational changes that *Ft*ClpP undergoes is similar to that previously described for other orthologues (Lee *et al.,* 2011). Work on *Ft*ClpP to find activators was not carried out due to time constraints and prioritisation of Ddl. Several enzymatic assays can be used for determining serine protease activity (Zhang, 2006). These include fluorescence the use of dyes (*e.g.* 7-amino-4-methylcoumarin and 7-amino-4(trifluoromethyl) coumarin) chemically

quenched with a peptide, fluorescence resonance energy transfer (FRET) and fluorescence polarisation using peptides labelled with a large size dye. The first is the most promising approach towards the development of a *Ft*ClpP high-throughput assay to screen compound libraries and search for activators.

PART SIX

DIHYDROFOLATE SYNTHASE:FOLYL-POLY-GLUTAMATE SYNTHASE

6.1 Introduction

Reduced folate derivates are essential nutrients in bacteria and eukaryotes. They are involved in one-carbon metabolism (Young *et al.,* 2008), for example the pathways for methionine, purine nucleotide and thymidylate biosynthesis (Kwon *et al.,* 2008). Folate, or vitamin B, can be synthesized *de novo* in most bacteria and plants but in mammals folate is generally obtained from extracellular sources and actively transported (Bermingham and Derrick, 2002). In bacteria, the folate biosynthesis pathway (Figure 6.1) involves several enzymes. The first four generate the intermediate 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (DHPPP) from GTP. The second part of the pathway uses DHPPP to synthetize active folate forms. Here, dihydropteroate synthase (DHPS, EC 2.5.1.15) generates dihydropteroate (H_2Pte). This is then used by dihydrofolate synthase (DHFS, EC 6.3.2.12) to form dihydrofolate (DHF). Dihydrofolate reductase (DHFR, EC 1.5.1.3) reduces DHF and generates tetrahydrofolate (THF) that can be poly-glutamated by folyl-poly-glutamate synthase (FPGS, EC 6.3.2.17). Folate can also be reduced by DHFR to form DHF, which can then proceed to poly-glutamation (Bognar and Shane, 1986; Sheng *et al.,* 2003). It has been shown that the poly-glutamated forms of the vitamin are essential in bacteria and eukaryotes, and that FPGS activity is present in both cases (Young *et al.,* 2008). Poly-glutamation provides a way to accumulate folic acid (Sheng *et al.,* 2002). Most folate-dependent proteins have higher affinity for the poly-glutamated forms (Young *et al.,* 2008). Indeed, even antifolate inhibitors can be poly-glutamated by FPGS activity (Sheng *et al.,* 2002). An example of this is methotrexate (MTX), a DHFR inhibitor

(Barnes *et al.,* 1999), where poly-glutamation facilitates retention inside the cell (Galivan *et al.,* 1985). The number of glutamates that are added to the folate derivatives is different between species, so that FPGS adds three glutamates in *E. coli* and up to nine in the case of *Lactobacillus casei* and mammals (Wang *et al.,* 2010).

In most bacteria, a single protein presents DHFS and FPGS activities. This is the bifunctional protein FolC. In fungi and some bacteria, these enzymes exist as separate proteins. In the case of mammals, DHFS activity is not present which suggests this might be a selective antibacterial target (Sheng *et al.,* 2008; Wang *et al.,* 2010). Fifteen FolC structures have been determined; two *E. coli* FolC in the presence of ADP or

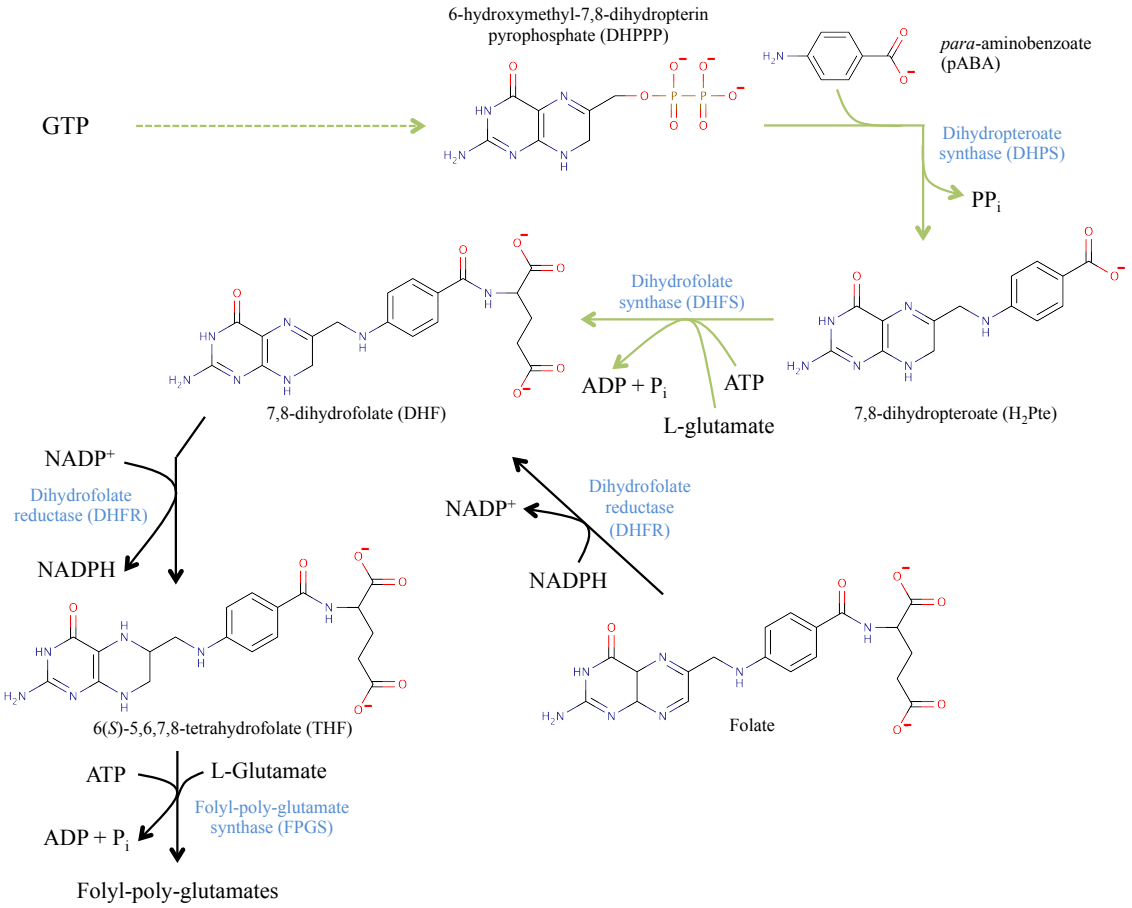


Figure 6.1: Schematic of the folyl-poly-glutamate biosynthesis pathway. Enzymes are colored in blue. Green arrows indicate the folate biosynthesis pathway in bacteria. Dashed green arrows represent the first four enzyme reactions that generate DHPPP. Black arrows mark the part of the pathway present in both prokaryotes and eukaryotes.

phosphorylated dihydropteroate+ADP (Mathieu *et al.,* 2005), two *M. tuberculosis* FolC structures with ADP or AMPPCP (Young *et al.,* 2008), seven *Lactobacillus casei* FolC apo-form mutants, in the presence of AMPPCP, AMPCP+5,10-methylene-6-hydrofolate, or pyrophosphate (Sun *et al.,* 1998; Sun *et al.,* 2001; Smith *et al.,* 2006), and four *Y. pestis* FolC, two without ligands (3NRS, 3N2A) and two in the presence of AMPPNP+Glu or AMPPNP (3QCZ, 3PYZ). FolC proteins contain two distinct α/β C-terminal and N-terminal domains. The ATP binding site is formed by residues from both domains and the substrate binding sites are located in the N-terminal domain. Two loops at the N-terminal domain undergo a conformational change when ATP and DHF bind and the overall FolC structure shows a more closed confirmation. From the structures, the co-factor pocket and active site have strong similarities with the Mur ligase superfamily, suggesting they undergo a similar reaction mechanism (Mathieu *et al.,* 2005). It has been proposed that both FolC reactions are carried out in an ordered sequential manner. First, ATP binds followed by the H_2Pte, or poly-glutamated folate, and then L-glutamate. The substrate is phosphorylated and this favours the incorporation of L-glutamate to form DHF. The first products to be released are ADP and phosphate followed by the ligation product (Sheng *et al.,* 2002; Mathieu *et al.,* 2005).

Inhibition of folyl-poly-glutamate biosynthesis has been achieved by targeting DHPS and DHFR but little has been done for inhibition of DHFS (Nzila, 2006; Wang *et al.,* 2010). Two inhibitors, mimicking the acyl-phosphate intermediates of both DHFS and FPGS, have been reported against *Plasmodium falciparum* FolC (*Pf*FolC) DHFS activity (Wang *et al.,* 2010). The studies showed selective inhibition towards DHFS is feasible and suggest *Pf*FolC could be used as a new drug target for malaria.

FolC was identified as a potential target for *B. anthracis* (*Ba*FolC), *B. pseudomallei* (*Bp*FolC) and *Y. pestis* (*Yp*FolC) as explained in Part Two. Initial studies for the assessment of FolC chemical validation involved the investigation of crystallisation conditions and enzymatic assays.

6.2 Experimental procedures

6.2.1 Recombinant protein production

Genes encoding *Ba*FolC (UniProt code Q81LD4, expected molecular mass 48.9 kDa), *Bp*FolC (Q63JM2, 47.1 kDa) and *Yp*FolC (Q0WDC2, 47.2 kDa) proteins were purchased from GeneScript and cloned into the pET15bTEV vector using the restriction enzymes NdeI (at the 3' DNA end) and BamHI (5' end). The inserts were 1311 (*Yp*FolC), 1316 (*Ba*FolC) and 1360 bp (*Bp*FolC). Expression vectors were sequenced and no mutations were found.

Heterologous expression using *E. coli* strain BL21(DE3) was obtained after the addition of 1 mM IPTG and cultures incubated at 20ºC overnight. The first affinity chromatography column (Ni-NTA) was carried out using buffer A1 (20 mM Tris-HCl pH 7.5, 20 mM NaCl and 10 mM $MgCl_2$) or A2 (20 mM HEPES pH 7.5, 200 mM NaCl and 10 mM $MgCl_2$) and buffer B (buffer A1/A2 with 800 M imidazole). Next, the His_6-tag was cleaved, during dialysis into buffer A1/A2, using 1 mg of TEV protease per 10 mg of protein overnight at 4ºC. A second Ni-NTA chromatography step was carried out to separate the His_6-tagged and non-His_6-tagged proteins. A high level of purity was confirmed with SDS-PAGE and MALDI-TOF MS analysis. The quaternary structure was studied with SEC, using a superdex 200 HiLoad 26/60 prep grade column (GE Heathcare), and native-PAGE (Invitrogen). Protein concentrations were determined

using the predicted ε(*Ba*FolC) = 44600 $M^{-1}cm^{-1}$, ε(*Bp*FolC) = 37150 $M^{-1}cm^{-1}$, ε(*Yp*FolC) = 52285 $M^{-1}cm^{-1}$ (values obtained from PROTPARAM, Gasteiger *et al.,* 2003) and the Lambert-Beer equation.

6.2.2 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF), also called Thermofluor, is a technique for measuring protein denaturation. The dye SYPRO Orange binds to unfolded proteins by interaction with hydrophobic areas. This generates an increase in the fluorescence of the dye (Lo *et al.,* 2004). In this project, DSF has been used to check FolC stability in the presence of different buffers to select a protein storage solution distinct to buffer A1. Protein unfolding was performed using increasing temperatures and the melting temperature (T_m) of each sample was determined.

The assay was carried out in a 96-well plate. The protein concentration used in each well was 4 μM, and SYPRO orange 1000 times diluted (stock purchased from Invitrogen). *Ba*FolC and *Yp*FolC samples were incubated for 30 minutes with 24 different buffers with different pH and salt concentrations. These involved Bicine, CAPS, CHES, HEPES, imidazole, MOPS, sodium-cacodylate, sodium-acetate, Tricine, Tris-HCl, and NaCl or KCl as the salt component. The increase in temperature and the fluorescence recordings were made using the real-time PCR system MX3005P from Stratagene.

6.2.3 Crystallisation and diffraction

*Ba*FolC, *Bp*FolC and *Yp*FolC were concentrated to 15 mg ml^{-1}. Prior to crystallisation TCEP at a final concentration of 1 mM was added. Initial crystallisation experiments were carried out using proteins in buffer A1 and samples were also incubated with 10

mM AMPPNP and L-glutamate. The screens tested were The PEGs Suite, The MPD Suite and The Classics Suite from Qiagen, and PGA, PROPLEX, JCSG plus and MIDAS from Molecular Dimensions. Drops were set up using 1:1 and 1:2 protein solution:reservoir ratios in final volumes of 0.2 and 0.3 µl. There were no positive hits from these screenings and only salt crystals grew in various conditions, and so the crystallisation trials were repeated using a different protein storage buffer. DSF gave information about the stability of the protein under various conditions; it identified buffers for which the protein was unstable and showed that higher NaCl concentration (0.5 M) provided an increase in *Yp*FolC stability. Due to the salt crystal growth in the first crystallisation experiments, the salt concentration was not increased and, since Mg^{2+} or Mn^{2+} is required for ATP hydrolysis, $MgCl_2$ was included. The resulting buffer contains 0.02 M HEPES, 0.2 M NaCl and 0.01 M of $MgCl_2$ (buffer A2). Based on the crystallisation conditions reported for PDB entries 3N2A and 3QCZ (*Yp*FolC in the presence of ADP and ANPPNP+Glu, respectively), a screen was designed containing 0.5-3 M sodium-formate, 20%-40% PEG 3350 and 0.1/0.2 M HEPES pH 7.0. Co-crystallisation of *Yp*FolC (in buffer A2) with 10 mM AMPPNP and L-glutamate was carried out and resulted in diffracting crystals. This crystallisation experiment was not performed with *Ba*FolC and *Bp*FolC proteins and might be considered for future work.

Data were collected using the in-house X-ray generator (Rigaku M007HF X-ray generator with a Saturn 944HG+ CCD detector). Data reduction and analysis were performed with XDS (Kabsch, 2009) and AIMLESS (Evans, 2011). The structure was solved using PhaserMR (McCoy *et al.,* 2007) and 3QCZ, a *Yp*FolC structure with AMPPNP and L-glutamate bound, as the search model. Ligands and waters were removed prior to molecular replacement. The space group of 3QCZ and the crystal of *Yp*FolC grown here is $P2_12_12_1$ and they have similar unit cell parameters (a=62.9

b=83.24 c=127.0 and a=62.0 b=78.0 c=131.0 Å, respectively). The Matthews coefficient determination using the CCP4i suite (Winn *et al.,* 2011) suggested the presence of one subunit in the asymmetric unit (V_m value of 3.4 $Å^3$ Da^{-1} and bulk solvent of 64%). A round of rigid body refinement using REFMAC (Murshudov *at al.* 2011) was carried out. It showed the presence of a positive density peak (5 σ) in the ATP binding site. The structure did not offer new information about *Yp*FolC so no further refinement was pursued.

6.2.4 Enzyme kinetics

Different enzymatic assays were tested with *Yp*FolC. Two coupled enzyme systems can be used (Figure 6.2); the formation of DHF can be measured using DHFR, and ATP hydrolysis to ADP can be recorded using the pyruvate kinase/lactate dehydrogenase (PK/LDH) system. An additional assay, BIOMOL Green, was used to measure the release of phosphate.

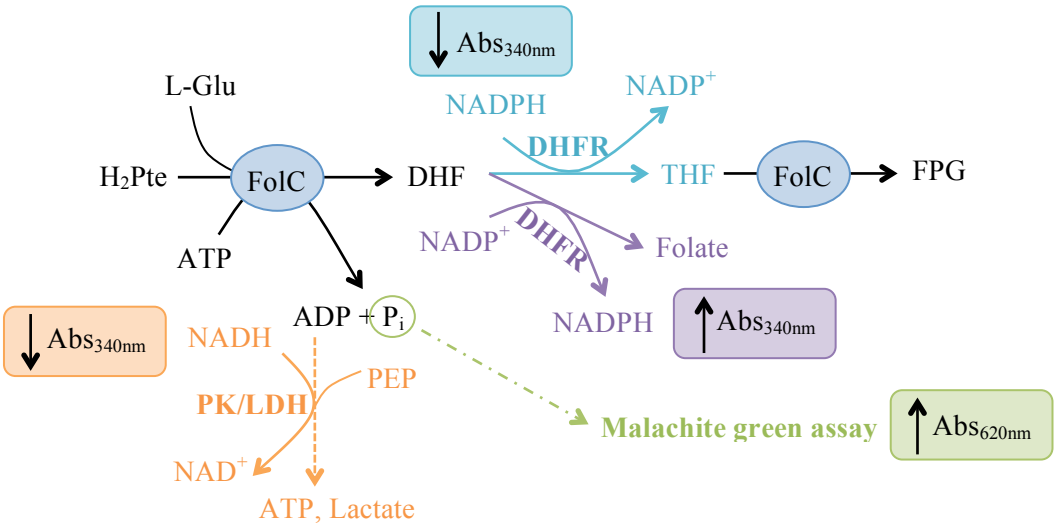


Figure 6.2: Scheme of assays tested for measuring FolC activity (blue circles). Coupling enzymes are shown in bold. The assay described by Bognar *et al.* (1985) is coloured in turquoise, and the modification of the assay in purple. The PK/LDH coupled assay is coloured in orange. The dashed orange arrow indicates a two-step reaction. The BIOMOL Green assay is marked in green and the dashed green arrow indicates non-enzymatic reactions. The expected effect in the absorbance at 340 or 620 nm is indicated with up (increase) and down (decrease) black arrows inside coloured squares. **H_2Pte** is dihydropteroate, **FPG** folyl-poly-glutamate, **DHR** dihydrofolate, **THF** tetrahydrofolate and **PEP** phospho-enol pyruvate.

6.2.4.1 Coupled spectrophotometric assays

DHFR coupled assay

Firstly, the activity of DHFR (human enzyme purchased from SIGMA) was checked. The reaction mixture contained 50 mM Tris-HCl pH 9, 100 μM NADPH, 25 μM DHF, 1 mM TCEP, 10 mM $MgCl_2$ and 50 mM KCl in a final volume of 1 ml. A negative control involved the reaction mixture without the addition of DHFR. Activity was measured at 5 and 10 mU ml^{-1} enzyme concentration for 10 minutes at 37ºC. The decrease in absorbance at 340 nm was recorded by a UV-2450 Shimadzu spectrophotometer. To determine *Yp*FolC activity, the assay mixture contained 50 mM Tris-HCl pH 9, 50 mM KCl, 1 mM TCEP, 10 mM $MgCl_2$, 100 μM NADPH, 5 mU ml^{-1} of DHFR, 25 μM H_2Pte, 5 mM ATP and 10 mM L-glutamate (1 ml final volume) (Bognar *et al.,* 1985). Tests at 37ºC for 30 minutes with different *Yp*FolC concentrations (from 10 to 200 μM) were carried out. Activity was also checked using an assay variation. Instead of using NADPH, 100 μM $NADP^+$ was used to promote the formation of folate and avoid the generation of THF, as this is also a substrate of *Yp*FolC (Figure 6.2). However, the folate formed might be transformed to DHF, thereby complicating the analysis, and this assay was used as an attempt to confirm whether *Yp*FolC was active. Same positive and negative controls were used in this assay variation. All controls showed the expected result.

PK/LDH coupled assay

The same procedure and assay buffer was used for the PK/LDH coupled system but instead of DHFR and NADPH, 5 μl ml^{-1} PK/LDH mixture from SIGMA and 70 μM NADH were used. Additionally, 1 mM phopho-enol pyruvate was added.

6.2.4.2 Malachite green assay

A malachite green assay was tested as a potential HTP assay, as wavelength distinct to 340 nm is required to avoid interference due to compounds. An enzyme titration experiment was carried out using the BIOMOL Green assay procedure shown in Part Four. The assay conditions included 50 mM Tris-HCl pH 9, 50 mM KCl, 1 mM TCEP, 10 mM $MgCl_2$, 100 μM H_2Pte, 500 μM ATP and 10 mM L-glutamate in a final volume of 30 μl in a clear bottom 384 well plate. The starting protein concentration used was 10 μM and serial 1/2 dilutions were made in every row. The absorbance at 620 nm was measured using the EnVision plate reader from PerkinElmer at 0, 2, 5, 10 and 20 minutes.

6.2.5 Tryptophan fluorescence

Tryptophan fluorescence was performed using a buffer containing 50 mM Tris-HCl pH 9, 50 mM KCl, 1 mM TCEP and 10 mM $MgCl_2$. Different concentrations of H_2Pte (0.3, 0.45, 0.6, 0.75, 0.9, 3, 4.5, 6, 7.5, 9, 30, 60, 90 μM) and L-glutamate (same concentrations) were tested in the presence and absence of 5 μM *Yp*FolC. Samples were exited at 280 nm and the emitted fluorescence recorded at 340 nm using the FLUOstar OPTIMA from BMG Labtech. The negative control was the buffer in the absence of protein or substrates, and the positive control was buffer and 5 μM *Yp*FolC. The assay was run in triplicate and a curve for measuring the fluorescence emission from H_2Pte was performed (same H_2Pte concentrations in the absence of *Yp*FolC). The corresponding H_2Pte signal was removed from the assay signal to determine the binding specific fluorescence. The K_d for H_2Pte was determined using the saturation binding equation from GraphPad (www.graphpad.com).

6.3 Results and discussion

6.3.1 Recombinant protein production

Heterologous expression was obtained with yields of 10, 80 and 90 mg per liter of culture of *Bp*FolC, *Ba*FolC and *Yp*FolC respectively. Purification was achieved with different buffers and the high levels of purity of the samples was confirmed with SDS-PAGE gels and MALDI-TOF MS analysis, which showed peaks corresponding to 49.03 (*Ba*FolC), 47.37 (*Bp*FolC) and 47.34 (*Yp*FolC) kDa being consistent with the expected protein masses.

The quaternary structure of FolC was investigated. All proteins eluted as monomers in SEC but the native-PAGE showed the presence of dimers and monomers. From the structures, monomers are expected but dimer formation has been suggested for DHFS activity (Bermingham and Derrick, 2002). Further analysis involving enzymatic assays will follow.

6.3.2 Differential scanning fluorimetry

The first crystallisation screens did not yield any crystal hits. This could be due to the buffer used for protein purification and storage. To investigate a more suitable protein buffer, DSF was carried out. The results for *Yp*FolC (Figure 6.3) showed good stability of the protein in various buffers and pH values. The T_m for *Yp*FolC in buffer A1 is 51ºC and no significant improvements in protein stability were observed. The protein had the highest T_m value in the solution containing 50 mM Tris-HCl and 500 mM NaCl (52ºC). DSF results also showed decreasing protein stability with imidazole (46ºC), MES (43ºC), sodium-acetate (37ºC) and sodium-cacodylate (46ºC). In the case of *Ba*FolC, no increase in the fluorescence signal was detected for reasons that are unclear.

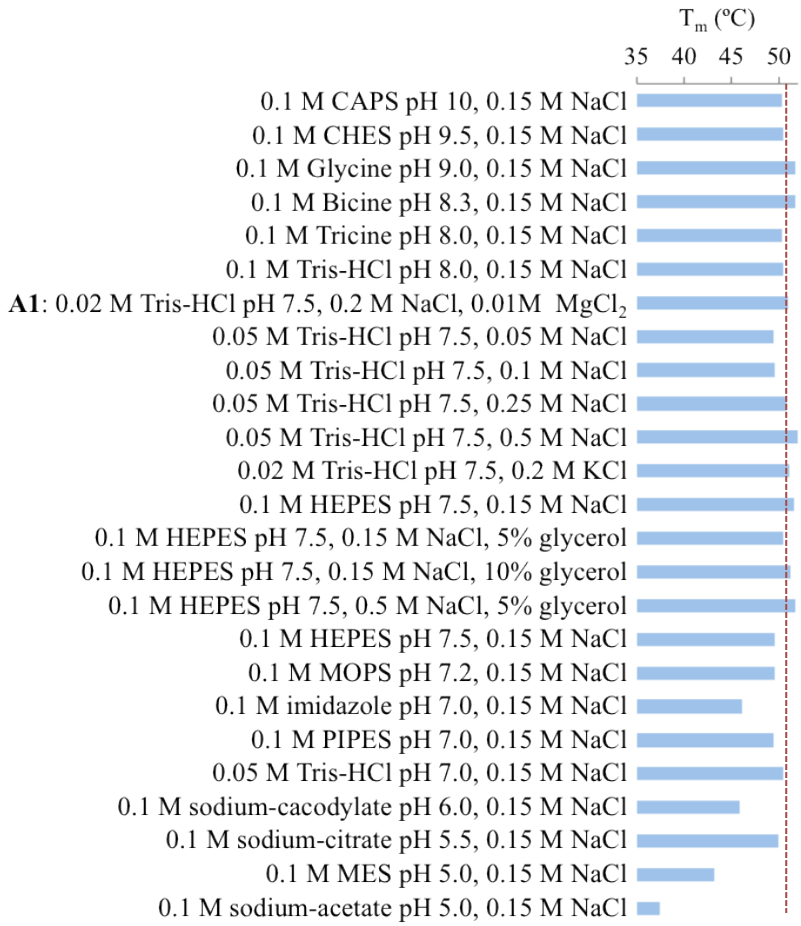


Figure 6.3: Results of *Yp*FolC DSF represented by the T_m (ºC). The dashed red line marks the T_m value of the protein in the presence of buffer A.

6.3.3 Crystallisation and diffraction

Four *Yp*FolC structures with different ligand combinations have been already solved by the Center for Structural Genomics of Infectious Diseases of Seattle, although no publications are available. The aim in this work for crystallisation of FolC was to find a condition that can be used for inhibitor:protein binding studies. Good quality crystals were obtained only in the case of *Yp*FolC (Figure 6.4A-B). These crystals diffracted to a resolution of 2.5 Å and the structure was determined. Positive density corresponding to AMPPNP was observed (Figure 6.4C). *Yp*FolC crystals could be further optimised for co-crystallisation and resolution improved at a synchrotron source. Due to the difficulties for determining the enzymatic activity of the protein (section 6.3.4), work in

crystal optimisation was stopped until an assay for inhibition validation could be developed, and some inhibitory hits found.

The crystallographic statistics are shown in Table 6.1.

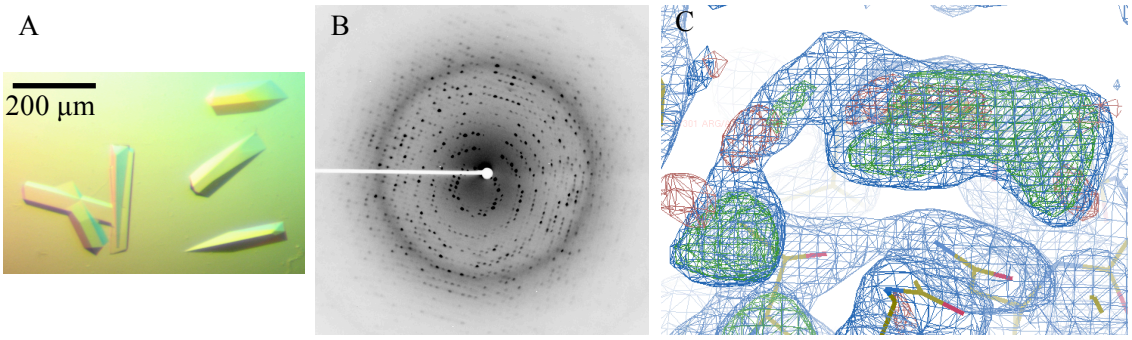


Figure 6.4: (A) *Yp*FolC optimised crystal in the presence of 10 mM AMPPNP, 10 mM L-Glu, 1 M sodium-formate, 0.1 M HEPES pH 7.5 and 20% (w/v) PEG 3350. (B) Diffraction pattern from collected data on the in-house X-ray generator. (C) Electron density (blue, 1.4 σ) and difference maps (in green, 3 σ) for AMPPNP at the ATP binding pocket of *Yp*FolC.

Structure	***Yp*FolC:AMPPNP**
Space group	$P2_12_12_1$
Wavelength (Å)	1.54178
Unit cell dimensions *a*, *b*, *c* (Å)	62.0, 78.0, 131.0
Resolution range[a] (Å)	45.5 - 2.5
No. reflections	79069
Unique reflections	22627
Completeness (%)	99.8 (99.5)
R_{merge}[b]	0.264 (0.450)
Redundancy	3.5 (3.3)
< I/σ(I)>	3.2 (2.1)
Wilson *B* (Å^2)	17.6
R_{work}[c]/R_{free}[d]	0.3660 / 0.3974*
Mosaic spread	0.5

Table 6.1: Crystallographic statistics. [a.] Values in parentheses refer to the highest resolution shell. [b.] R_{merge} = $\sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/ \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. [c.] R_{work} = $\sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. [d.] R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations. (*) Refinement not complete; values after one cycle of restraint refinement.

6.3.4 Enzyme kinetics

As indicated in the introduction, the next enzyme in folyl-poly-glutamate biosynthesis after the first FolC reaction is DHFR. This enzyme oxidizes NADPH which can be measured as a decrease in the absorbance at 340 nm. Bognar *et al.* (1985) published an assay for determining enzyme kinetics parameters of *L. casei* and *Corinebacterium sp.* FolC using DHFR as a coupled enzyme. In this work, it was not possible to observe FolC activity using this assay. DHFR might not be useful for measuring FolC activity since the product THF is also a FolC substrate. Even using the variation of Bognar *et al.* (1985) assay, DHFR could form THF using DHF synthesised by FolC compromising the assay performance. To avoid using DHFR, another coupled assay involving PK and LDH was tested but no activity was detected, although binding of the substrate, H_2Pte, to FolC was confirmed using tryptophan fluorescence (see section 6.3.5). Other groups have carried out enzyme kinetic studies for various FolC orthologues using radioactivity (Murata *et al.,* 2000; Garrow *et al.,* 2004; Wang *et al.,* 2010) or HPLC (high performance liquid chromatography, Sybesma *et al.,* 2003; Mathieu *et al.,* 2005) instead of the coupled assay previously described. This may be indicative of an issue with the assay, probably due to interference of the substrate at the measured wavelength and/or the DHFR product. As a last check, the BIOMOL Green assay was used to measure the phosphate release and *Yp*FolC activity was detected (Figure 6.5). Further experiments are required to determine the K_m values of H_2Pte, L-glutamate and ATP and to study the reaction mechanism.

The result of the enzyme titration experiment is shown in Figure 6.5. Once the K_m values have been determined, this could be used for developing a high-throughput assay.

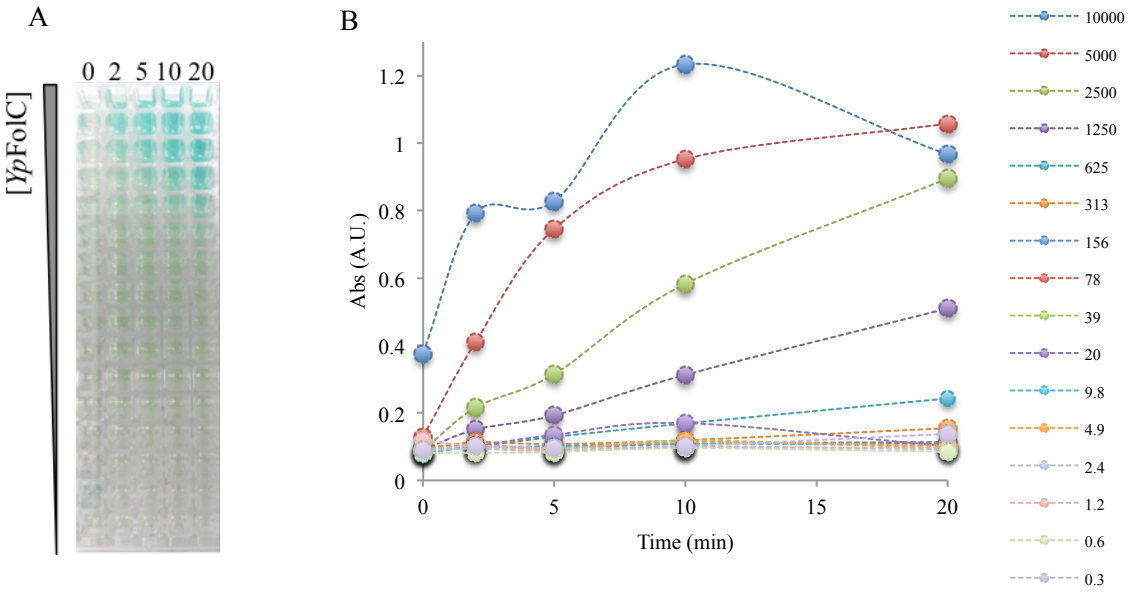


Figure 6.5: (A) Result from *Yp*FolC enzyme titration experiment using the BIOMOL Green assay. (B) Plot representation of *Yp*FolC results in A. Absorbance at 620 nm at different time points (min) and *Yp*FolC concentrations (nM).

6.3.5 Tryptophan fluorescence

Tryptophan fluorescence was used to measure the binding constant (K_d) of H_2Pte and L-glutamate. In the first case, the results showed a K_d of 24 µM (Figure 6.6). For L-glutamate, variations in the tryptophan fluorescence were not observed. Fluorescence changes may be due to the presence of two tryptophan residues in a loop that changes its conformation when ATP and H_2Pte bind. L-glutamate binding might not be involved in the re-organization of this loop, and so has no effect on tryptophan fluorescence.

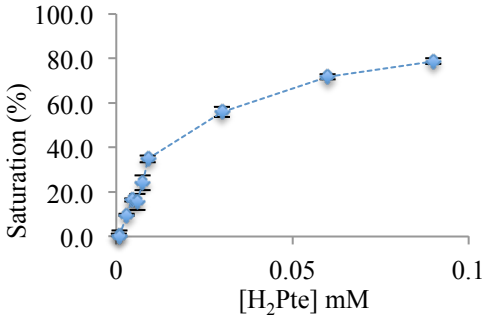


Figure 6.6: *Yp*FolC tryptophan fluorescence result. The plot represents the percentage of saturation against the substrate concentration.

The results confirmed H_2Pte binding. From the previously described sequential ordered mechanism (Mathieu *et al.,* 2005) ATP is the first compound to bind FolC. In this experiment, binding of H_2Pte to *Yp*FolC is observed in the absence of ATP which might suggest the binding of the substrates is not necessarily ordered. This has to be confirmed with enzyme activity characterisation.

6.4 Conclusions

Bifunctional FolC is a potential selective antibacterial drug target as mammals lack DHFS activity. *Yp*FolC crystals have been obtained and optimisation can be conducted for ligand co-crystallisation. Other techniques for ligand binding studies are BLI or SPR (Surface Plasmon Resonance), but it would be necessary to test difference ligand bound states of the enzyme. The search for an enzymatic assay has been pursued and a new DHFS assay based on the detection of released phosphate was found. Future work should focus on the optimisation of this assay to perform biochemical characterisation and develop a HTP assay to support a compound screening campaign.

PART SEVEN

CONCLUSIONS

This project aimed to evaluate and carry out initial validation of several potential antibacterial drug targets. Ten different enzymes in the biowarfare agents *B. anthracis*, *Y. pestis*, *B. pseudomallei* and *F. tularensis* were selected on the basis of the genome-scale transposon knock-out sequencing data and previous work from the Dstl and Hunter labs. Structural and biochemical studies were initiated on four of these, KynB, Ddl, FolC and ClpP, and the targets were then prioritised based on the empirical results.

- Structures were determined for three bacterial KynB examples. The structures show a distinct amidase fold and binuclear Zn^{2+} catalytic site, unrelated to the eukaryote kynurenine formamidase. This indicates that specificity towards bacteria might be feasible. However, genetic analysis showed no essentiality and work was directed to other targets, highlighting the importance of the genetic validation of potential targets. Nevertheless, should KynB be proven essential for other organisms, studies carried out during this project give information and tools for further validation. There is some evidence to suggest KynB is involved in virulence of *B. pseudomallei*, so inhibitors for KynB might be used in combinatorial therapies.
- Research on *B. pseudomallei* Ddl provided structural and kinetic data. An optimised high-throughput (HTP) assay was developed and used to screen compound libraries. Despite testing a large number of compounds, no inhibitors were found suggesting other approaches, such as structure-based rational drug design, might be used for assessing this challenging target. In this case, inhibitors have been developed by others since 1988 but no information about the follow up has been described suggesting problems with the characterised

compounds. The optimised high-throughput assay can be used to screen new compound libraries that might be developed in the future.

- Well ordered crystals have been obtained for *Y. pestis* FolC and can be further optimised to develop ligand-bound studies. Problems were encountered towards the development of an enzymatic assay for *Y. pestis* FolC, and a new assay has been found for measuring dihydrofolate synthase activity. Future work could be focused on the assay optimisation to characterise the enzyme and perform HTP compound library screens.
- Pure *F. tularensis* ClpP was obtained and used for crystallisation. Two structures were determined showing the compressed and the open conformations. These data can be used for the development of compounds targeting this protein, or once activators are found, characterising the ligand binding modes.

The project has generated data on the assessment of several potential drug targets. Complete early-stage chemical validation was not achieved. Nevertheless, only Ddl could be suggested as a target in *B. pseudomallei* taking into account the need for new inhibitors.

Different therapies are being developed and improved to combat (multi)drug resistant bacteria. These include the use of conventional antibiotics, phage therapy (for which several clinical trials are being held for *E. coli*, *P. aeruginosa* and multidrug resistance *S. aureus* infections, Kingwell, 2015) and RNA interference (Dyawanapelly *et al.,* 2014). Despite the technology and the information available for the development of antibacterial drugs, the rate of success in drug discovery remains low and the process should be improved. On one hand, the report of negative results from protein

characterisation, target assessment and inhibitors follow up projects, as performed in this thesis, could be key for improving the rate of success by designing more appropriate strategies to validate targets. On the other hand, it has been shown that antibiotics and neuroactive compounds have different chemical properties (O'Shea and Moser, 2007; Payne *et al.,* 2007). Regarding this, and the low success rate of HTP (Payne *et al.,* 2007; Gwynn *et al.,* 2010; Chopra 2013; Tommasi *et al.,* 2015), it seems libraries still need to be improved. In this context, the re-evaluation of current compound libraries to become more antibiotic-specific might help towards a more effective antibiotic search. Therefore, it is reasonable to think that bacteria type and niche will require compounds with distinct chemical properties. For example, the cell membrane and peptidoglycan layers are different between Gram-positive and Gram-negative bacteria, the presence of external capsules, formation of biofilms, the location of the infection or if bacterial cell division occurs inside another cell-type (*e.g. B. pseudomallei*, *F. tularensis* and *Y. pestis*, see Part Two), are also indicative that drugs may require different chemical properties. Screening current libraries in a phenotypic manner against different bacteria types could identify compounds that could overcame barriers and those that do not. This could help to create a collection of chemical properties specific to each bacterial type/species. This information could then be used to generate new bacteria-specific libraries for either phenotypic or single-target screens. Additionally, the identification of biologically active chemical scaffolds could be used for structure-based drug design to improve compound permeability. Combinatorial therapies involving the use of several drugs should be further studied as this may decrease the probability of drug resistance development. Additionally, drugs already in the market could be combined with newly developed compounds (*e.g.* efflux pumps inhibitors) to re-purpose existing drugs. Finally, the search of potential essential genes and candidate selection could be accompanied by performing genome-scale transposon

knock-outs in the presence of antibiotics to identify genes involved in drug resistance or that are essential under drug stress conditions.

Appendix A

Results from TraDIS and the target selection criteria analysis for the 75 considered candidates.

Table A.1: Target selection criteria and traffic light score. Bacteria abbreviations are: *Bacillus anthracis* (*Ba*), *Burkholderia pseudomallei* (*Bp*), *Bombyx mori* (*Bm*), *Bos taurus* (*Bt*), *Escherichia coli* (*Ec*), *Helicobacter pylori* (*Hp*), *Homo sapiens* (*Hs*), *Micrococcus luteus* (*Ml*), *Mycobacterium tuberculosis* (*Mt*), *Neisseria gonorrhoeae* (*Ng*), *Pseudomonas aeruginosa* (*Pa*), *Plasmodium falciparum* (*Pf*), *Rattus norveqieus* (*Rn*), *Saccharomyces cerevisiae* (*Sc*), *Salmonella enterica* (*Se*), *Streptococcus pneumoniae* (*Sp*), and *Yersinia pestis* (*Yp*). E indicates the gene is essential for cell growth/viability, P probable essential and N non-essential. When a human homologue was present, the percentage of sequence identity was included in the table. In the ChEMBL column the first values corresponds to the number of ChEMBL homologues and the second the number of ChEMBL targets. The Size column indicates the number of amino acids (AA) for the *P. aeruginosa* protein. Targets selected for further study are marked in blue. XtalPred score ranges from 1 (optimal crystallisation) to 5 (challenging crystallisation). See Part Two for traffic light score definitions.

Yp gene name	Gene product	EC	Essentiality		Human homolgue	Druggability			Tractability as a target				
			Pa	DEG		QED	ChEMBL	Drug-like compounds	Size (AA)	Location	Assay	Structures (resolution)	XtalPred
accA	Acetyl-CoA carboxylase carboxyltransferase subunit alpha	6.4.1.2	E	E	No		0 / 0	Inhibitors (nM). (Freiberg *et al.,* 2004)	316	Cytoplasm	HPLC and radioactivity. HTP assay: Spectrophotometry (Soriano *et al.,* 2005; Santoro *et al.,* 2006)	9 structures, dimer, apo and ligand-bound (1.5 - 3.0 Å)	3
adk	Adenylate kinase	2.7.4.3	P	E	Yes, 50%	0.03	11 / 0	Inhibitors *Ec*: AMP, ATP analogues/derivatives (nM). (BRENDA database)	215	Cytoplasm	Spectrophotometric coupled assay. (Saint *et al.,* 1986)	66 structures, monomer, apo and ligand-bound, *Bp*, (1.2 - 3 Å)	1
alaS	Alanyl-tRNA synthetase	6.1.1.7	N	E	Yes, 40%		2 / 0	Toxic inhibitors *Ec* (BRENDA database)	874	Cytoplasm	Radioactivity, Single turnover kinetics (Zhang *et al,* 2006)	Monomer, apo and ligand-bound (2.3 - 3.0 Å)	5
argS	Arginyl-tRNA synthetase	6.1.1.19	P	E	Yes, 37%	0.07	1 / 0	None	587	Cytoplasm	Radioactivity (Kiga *et al.,* 2001)	Monomer, apo and ligand-bound (2.2 - 2.6 Å)	3
aspS	Aspartyl-tRNA synthetase	6.1.1.12	N	E	Yes, 40%	0.4	2 / 0	Inhibitor (nM) for *Ft* and *Ec* (Metlitskaya *et al.,* 2006). Adenilate analogues inhibit mitochondrial human homologue (Messmer *et al.,* 2009)	591	Cytoplasm	Radioactivity (Thompson *et al.,* 2007)	Monomer, apo and ligand-bound (1.5 - 3.3 Å)	3
cca	tRNA nucleotidyl transferae/2'3'-cyclic phosphodiesterase/ 2'nucleotidase	2.7.7.72 3.1.3.- 3.1.4.-	N	E	No		0 / 0	None	410	Cytoplasm	Multifunctional protein: CCA-tRNA nucleotidyltransferase, 2'-nucleotidase, 2',3'-cyclic phosphodiesterase, Phosphatase. Radioactivity (Yue *et al.,* 1996)	7 structures, monomer and dimer, (2.37 - 3.5 Å)	2
cysS	Cysteinyl-tRNA synthetase	6.1.1.16	E	E	Yes, 34%	0.23	2 / 0	None	460	Cytoplasm	Radioactivity (Ruan *et al.,* 2004)	5 structures, monomer, apo and ligand-bound, (2.3 - 2.55 Å)	3
dapA	Dihydrodipicolinate synthase	4.3.3.7	E	E	Yes, 30%	0.52	6 / 0	Some inhibitors (mM) for DapA *Ec* (BRENDA database)	292		Spectrophotometric coupled assay (Dobson *et al.,* 2004)	77 structures, dimer, apo and ligand-bound, *Ba* (1.2 - 3.0 Å)	1
dapE	Succinyl-diaminopimelate desuccinylase	3.5.1.18	N	E	No	0.54	1 / 0	Inhibitors *Ec* (nM) (Gillner *et al.,* 2009; Uda and Creus, 2011)	383	Cytoplasm	Spectrophotometric assay (Bienvenue *et al.,* 2003)	11 structures, dimer, apo and ligand-bound, (1.5 - 3.0 Å)	2
ddl	D-alanine--D-alanine ligase	6.3.2.4	N	E	No		2 / 0	Inhibitors *Ec*, *Hp* (nM): Quercetin and apigenin (Wu *et al.,* 2008)	319	Cytoplasm	Spectrophotometric coupled assay. (Wu *et al.,* 2008)	29 structures, dimer, apo and ligand-bound, *Ba*, *Bp*, *Yp* (1.5 - 3.0 Å)	2
dfp	Phosphopantothenoylcysteine decarboxylase/phosphopantothenate synthase	4.1.1.36 6.3.2.5	E	E	No		0 / 0	None	402	Cytoplasm	Phosphopantothenoylcysteine decarboxylase. Phosphopantothenate--cysteine ligase. Radioactivity (Strauss *et al.,* 2001)	None	2
fabG	3-ketoacyl-ACP reductase	1.1.1.100	E	E	Yes, 40%	0.52	25 / 0	Inhibitor (nM) for *Ec* (Sun *et al.,* 2008)	247	Cytoplasm	Spectrophotometric assay (Sun *et al.,* 2007)	70 structures, tetramer, *Bp*, *Ba* (1.5 -3.0 Å)	1
fabH	3-oxoacyl-[acyl-carrier-protein] synthase	2.3.1.180	N	E	No	0.73	6 / 0	Inhibitors (nM) for *Ec* (Zhang *et al.,* 2011; Lee *et al.,* 2009)	330	Cytoplasm	Radioactivity (He and Reynolds, 2002)	58 structures, dimer, *Yp*, *Bp* (1.5-3.0Å)	2
fabZ	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase	4.2.1.59	E	E	No	0.53	1 / 0	Inbitors of *Mt* (He *et al.,* 2009), IC50 2 µM	146	Cytoplasm	Spectrophotometric assay (He *et al.,* 2009)	19 structures, 6-mer, apo and ligand-bound (2.0 - 3.0Å)	2
folC	Bifuntional protein: folylpolyglutamate synthetase	6.3.2.17 6.3.2.12	E	E	Yes, 30%	0.27	2 / 0	Two inhibitors for malaria (Wang *et. al* 2010)	429	Cytoplasm	Spectrophotometric assay (Bognar *et al.,* 1985)	15 structures, monomer, apo and ligand-bound, *Yp* (1.5 - 2.5 Å)	2
folD	5,10-methylene-tetrahydrofolate dehydrogenase/cyclohydrolase	1.5.1.5 3.5.4.9	E	E	Yes, 43%		2 / 0	None	284	Cytoplasm	Methylenetetrahydrofolate dehydrogenase: Spectrophotometric assay (Ramasastri and Blakley, 1962). Methenyltetrahydrofolate cyclohydrolase: Spectrophotometric assay (Schmidt *et al.,* 2000)	1 structure dimer, *Ec* (2.6 Å)	1
folP	Dihydropterate synthase	2.5.1.15	E	E	No	0.81	4 / 0	Inhibitor (nM) for *Ba* (Valderas *et al.,* 2008)	283	Cytoplasm	Radioactivity (Roland *et al.,* 1979), Spectrophotometric coupled assay (Valderas *et al.,* 2008)	45 structures, monomer, apo and ligand-bound, *Yp*, *Ba, Ft* (1.8 - 2.2 Å)	1

Yp gene name	Gene product	EC	Essentiality		Human homolgue	Druggability			Tractability as a target				
			Pa	DEG		QED	ChEMBL	Drug-like compounds	Size (AA)	Location	Assay	Structures (resolution)	XtalPred
ftsI	Peptidoglycan synthetase	2.4.1.129	N	E	No	0.65	13 / 0	None	579	Inner membrane	Radioactivity (Terrak *et al.*, 2008)	16 structures, monomer-multimer, apo and ligand-bound (1.5 - 2.5 Å)	5
ftsZ	Cell division protein FtsZ	3.6.5.6	E	E	No	0.57	5 / 1.	Inhibitors for *Bt* (BRENDA database)	394	Cytoplasm	Radioactivity (Lu *et al.*, 1999)	63 structures, monomer-multimer, apo and ligand bound (1.2 - 4.0 Å)	5
fusA	Elongation factor G	3.6.5.3	N	E	Yes, 43%		2 / 0	There are 2 froms of fusA in *Pa* (84% homology). Inhibitor for *Pf* (Gupta *et al.*, 2013), *Ec* (Hausner *et al.*, 1988)	706 702	Cytoplasm	Radioactivity (Harmark *et al.*, 1992)	7 structures, monomer and complexes, (2.5 - 3.5 Å)	5
glmM	Phosphoglucosamine mutase	5.4.2.10	N	E	No		0 / 0	None	445	Cytoplasm	Radioactivity coupled assay and spectrophotometric coupled assay (Jolly *et al.*, 1999)	2 structures, *Ft*, *Ba* (2.3 - 2.7 Å)	1
glmU	N-acetylglucosamine-1-phosphate uridyltransferase	2.7.7.23 2.3.1.157	E	E	No	0.55	2 / 0	Inhibitors for different *sp*. (Green *et al.*, 2012)	454	Cytoplasm	UDP-N-acetylglucosamine pyrophosphorylase: Radioactivity (Bullik *et al.*, 2000). Glucosamine-1-phosphate N-acetyltransferase: Ellman's reagent, spectrophotometric assay (Buurman *et al.*, 2011)	45 structures, trimer, *Yp* (2.0 - 3.4 Å)	2
glnA	Glutamine synthetase	6.3.1.2	E	E	No	0.36	3 / 0	Inhibitor examples for different *sp.* (BRENDA)	469	Cytoplasm	Direct or coupled spectrophotometric assay (Pearson *et al.*, 2005)	18 structures, 12-mer, (1.5 - 3.5 Å)	3
glnS	Glutaminyl-tRNA synthetase	6.1.1.18	E	E	Yes, 48%	0.2	2 / 0	Inhibitors for *Ec* (nM) (Bernier *et al.*, 2000)	556	Cytoplasm	Radioactivity (Liu *et al.*, 1998)	21 structures, dimers and tRNA-complexed, (1.5 - 3.5 Å)	4
gltA	Citrate synthase	2.3.3.1	N	E	Yes, 26%	0.57	1 / 0	None	428	Cytoplasm	Spectrophotometric assay (Pereira *et al.*, 1994)	14 structures, 6-mer, apo and ligand-bound, *Ft* (1.5 - 3.0 Å)	1
gltX	Glutamyl-tRNA synthetase	6.1.1.17	E	E	Yes, 35%	0.64	2 / 0	Inhibitor for *Ec* (nM) (Balg *et al.*, 2007)	494	Cytoplasm	Radioactivity (Lapointe and Söll, 1972)	4 structures, dimer and monomer, apo and ligand-bound, *Burkholderia sp.* (1.5 - 3.0 Å)	3
glyS	Glycyl-tRNA synthetase subunit beta	6.1.1.14	E	E	No		0 / 0	Inhibitors (nM) for *Bm* (Dignam *et al.*, 2003)	684	Cytoplasm	Radioactivity (Johanson *et al.*, 2003)	3 structures, monomer, apo and ligand-bound (2.3 - 3.4 Å)	5
groEL	Molecular chaperone GroEL	3.6.1.3	P	E	Yes, 51%		2 / 0	None	547	Cytoplasm	ATPase. Radioactivity (Vineyard *et al.*, 2006)	47 structures, multi-mer, *Yp* (1.5 - 4.5 Å)	3
hemC	Porphobilinogen deaminase	2.5.1.61	E	E	Yes, 43%		0 / 0	None	313	Cytoplasm	Flourescence assay (Schneider-Yin *et al.*, 2008)	2 structures, monomer, *Bacillus sp.* (1.5 - 2.0 Å)	1
hemL	Glutamate-1-semialdehyde aminotransferase	5.4.3.8	P	E	Yes, 29%		2 / 0	Inhibitors for *Synechococcus sp.* and others (BRENDA)	427	Cytoplasm	Calorimetry (Smith *et al.*, 1991)	9 structures, dimer, *Ba, Yp* (1.5 - 3.0 Å)	1
hisS	Histidyl-tRNA synthetase	6.1.1.21	N	E	No	0.18	1 / 0	Inhibitors for *Ec* and *Se* (BRENDA)	429	Cytoplasm	Radioactivity (Rosen *et al.*, 2006)	8 structures, monomer and tRNA-complexed (2.0 - 3.0 Å)	2
ileS	Isoleucine--tRNA ligase 1	6.1.1.5	E	E	Yes, 36%	0.41	6 / 0	Inhibitors for *Ec* and others (BRENDA)	943	Cytoplasm	Radioactivity (Dulic *et al.*, 2010)	3 structures, monomer (1.5 - 2.5 Å)	5
ispB	Octaprenyl diphosphate synthase	2.5.1.90	N	E	Yes, 28%	0.34	7 / 0	Inhibitor for *Pf* (nM) (Tonhosolo *et al.*, 2005)	322	Cytoplasm	Radioactivity (Guo *et al.*, 2004).	3 structures, dimer (1.5 - 2.5 Å)	2
ispD	2-C-methyl-D-erythritol 4-phosphate cytidylytransferase	2.7.7.60	P	E	No		1 / 0	None	234	Cytoplasm	Radioactivity (Richard *et al.*, 2004). HTP assay: coupled spectrophotometric assay (Illarionova *et al.*, 2006)	24 structures, dimer and 6-mer, apo and ligand-bound (1.5 - 3.0 Å)	5
ispE (ipk)	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	2.7.1.148	E	E	No		1 / 0	Inhibitors for *Ec* (nM) (BRENDA; Hirsch *et al.*, 2007)	282	Cytoplasm	Spectrophotometric assay (Sgraja *et al.*, 2008). HTP assay: coupled spectrophotometric assay (Illarionova *et al.*, 2006)	16 structures, monomer, apo and ligand-bound (1.5 - 2.5 Å)	2
ispG	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	1.17.7.1	E	E	No		0 / 0	None	371	Cytoplasm	Radioactivity (Zepeck *et al.*, 2005)	13 structures, dimer, apo and ligand-bound, *Ba* (1.3 - 3.0 Å)	1

Yp gene name	Gene product	EC	Essentiality		Human homolgue	Druggability			Tractability as a target				
			Pa	DEG		QED	ChEMBL	Drug-like compounds	Size (AA)	Location	Assay	Structures (resolution)	XtalPred
lepB	Signal peptidase I	3.4.21.89	N	E	No	0.52	1 / 0	Inhibitors for *Ec* and *Staphylococcus* (nM) (BRENDA).	284	Inner membrane	SDS-PAGE (PONA) (Carlos *et al.*, 2000)	7 structures, dimer, apo and ligand-bound (1.5 - 3.2 Å)	5
leuS	Leucyl-tRNA synthetase	6.1.1.4	N	E	Yes, 36%	0.41	7 / 0	Inhibitors for *Ec* (no K_i values) (BRENDA)	873	Cytoplasm	Radioactivity (Xu *et al.*, 2004)	5 structures, monomer, apo and ligand-bound (1.5 - 3.0 Å)	5
lipA	Lipoyl synthase	3.1.1.3	N	E	No		0 / 0	Inhibitors for *Ec*, *Pa* and other *sp.* (no K_i values) (BRENDA)	311	Outer membrane	Assay described for PseA lipase: spectrophotometric assay. (Kilcawley *et al.*, 2002)	22 structures, monomer, apo and ligand-bound (1.3 - 2.5 Å)	5
lpxA	UDP-N-acetylglucosamine acyltransferase	2.3.1.129	E	E	No		0 / 0	Inhibitors for *Ec* (Shapiro *et al.*, 2013) (Anderson *et al.*, 1993)	258	Cytoplasm	Radioactivity (Anderson *et al.*, 1993). HTP assay: fluorescence polarization assay (Shapiro *et al.*, 2013)	17 structures, trimer, apo and ligand-bound (1.4 - 3.0 Å)	3
lpxB	Lipid-A-disaccharide synthase	2.4.1.182	E	E	No		0 / 0	Inhibitor for *Ec*: Octyl-beta-D-glucoside (Ray *et al.*, 1984)	378	Cytoplasm	Radioactivity (Radika and Raetz *et al.*, 1988)	None	4
lpxC	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	3.5.1.108	E	E	No	0.54	3 / 1.	Inhibitor examples for *Pa.* (McAllister *et al.*, 2012) and for *Ec* (BRENDA)	303	Cytoplasm	Radioactivity (Liang *et al.*, 2013; Hernick *et al.*, 2010)	38 structures, monomer, apo and ligand-bound (1.3 - 3.1 Å)	1
lpxD	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	2.3.1.191	N	E	No		0 / 0	Inhibitors for *Ec* (nM) (BRENDA)	353	Cytoplasm	Radioactivity (Bartling anad Raezt, 2008)	7 structures, trimer, apo and complexed with other proteins (1.3 - 3.1 Å)	3
lpxK	Tetraacyldisaccharide 4'-kinase	2.7.1.130	E	E	No		0 / 0	None	332	Cytoplasm	Radioactivity (Emptage *et al.*, 2012)	None	2
lspA	Lipoprotein signal peptidase	3.4.23.36	E	E	No		0 / 0	Inhibitors for *Ec* (no K_i values) (BRENDA)	169	Inner membrane	Radioactivity (Sankaran and Wu, 1995)	None	5
lysS	Lysine--tRNA lig	6.1.1.6	E	E	Yes, 42%	0.4	1 / 0	Inhibitors for *Ec* and *Sc* (nM) (BRENDA)	501	Cytoplasm	Radioactivity (Ataide and Ibba *et al.*, 2004)	16 structures, dimer, apo and ligand-bound (1.5 - 3.1 Å)	5
metK	S-adenosylmethionine synthetase	2.5.1.6	N	E	Yes, 60%	0.25	2 / 0	Inhibitors for *Ec and Sc* (mM-nM) (BRENDA)	396	Cytoplasm	Radioactivity (Taylor and Markham, 2003)	14 structures, 4-mer, apo and ligand-bound (1.5 - 3.2 Å)	3
mraY	Phospho-N-acetylmuramoyl-pentapeptide transferase	2.7.8.13	E	E	No	0.2	4 / 0	Inhibitor for *Ec* (nM) (Brandish *et al.*, 1996) and others (BRENDA)	360	Inner membrane	Radioactivity (Brandish *et al.*, 1996)	1 structure, dimer, apo (3.3 Å)	5
msbA	Lipid A export ATP-binding/permease protein	3.6.3.39	E	E	Yes, 38%	0.8	52 / 0	Inhibitors for *Ec* (no K_i values) (BRENDA)	603	Inner membrane	Calorimetry (Ghanei *et al.*, 2007)	8 structures, dimer, apo and ligand-bound, (2.5 - 5.5 Å)	5
murD	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	6.3.2.9	P	E	No		6 / 0	Inhibitors (mM) for *Ec* (BRENDA)	448	Cytoplasm	Radioactivity (Bouhss *et al.*, 1999)	23 structures, monomer, apo and ligand-bound (1.4 - 3.0 Å)	1
murE	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase	6.3.2.13	E	E	No		6 / 0	Inhibitors (mM) for *Ec*, *Mt* and *Pa* (BRENDA)	487	Cytoplasm	Radioactivity (Mengin-Lecreulx *et al.*, 1994)	16 structures, monomer, apo and ligand-bound (1.5 - 3.0 Å)	2
murF	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	6.3.2.10	E	E	No		6 / 0	Inhibitor for *Sp* (Turk *et al.*, 2013)	458	Cytoplasm	Spectrophotometric coupled assay (Turk *et al.*, 2013)	12 structures, monomer, apo and ligand-bound (1.5 - 3.0 Å)	1

Yp gene name	Gene product	EC	Essentiality		Human homolgue	Druggability			Tractability as a target				
			Pa	DEG		QED	ChEMBL	Drug-like compounds	Size (AA)	Location	Assay	Structures (resolution)	XtalPred
murG	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	2.4.1.227	E	E	No		0 / 0	Inhibitors (nM) for *Ec* (BRENDA; Crouvoisier *et al.*, 2007)	357	Inner membrane	Fluorescence coupled enzyme assay (Liu *et al.*, 2003). Radioactivity (Crouvoisier *et al.*, 2007)	6 structures, monomer, apo and ligand-bound (1.5 - 3.0 Å)	3
nuoH	NADH-quinone oxidoreductase subunit H	1.6.99.5	N	E	Yes, 41%	0.69	2 / 0	Some inhibitors (nM) for *Ec* for NADH-quinone oxidoreductase (complex I) activity (BRENDA)	331	Inner membrane	Spectrophotometric assay (Jensen *et al.*, 2002) for NADH-quinone oxidoreductase (complex I) activity	None	5
nuoI	NADH dehydrogenase subunit I		N	E	Yes, 41%	0.12	1 / 0		182	Cytoplasm		16 structures, 6-mer, apo and ligand-bound (1.5 - 4.5 Å)	4
nuoJ	NADH-quinone oxidoreductase subunit J		N	E	No		0 / 0		166	Inner membrane		None	5
obgE	GTPase ObgE	3.6.5.2	E	E	Yes, 42%		0 / 0	There are inhibitors examples for other monomeric GTPases at BRENDA (EC 3.6.5.2)	406	Cytoplasm	monomeric GTPase. There are assays (radioactivity) for other monomeric GTPases at BRENDA (EC 3.6.5.2)	None	5
pheS	Phenylalanyl-tRNA synthetase subunit alpha	6.1.1.20	E	E	Yes, 31%	0.6	6 / 0	Inhibitors for *Sp* (BRENDA; Montgomery *et al.*, 2009)	338	Cytoplasm	Radioactivity. Both subunits are needed for enzyme activity (Roy *et al.*, 2004)	11 structures, 4-mer, complex (2.5 - 3.0 Å)	2
pheT	Phenylalanyl-tRNA synthetase subunit beta		E	E	No		1 / 0		792	Cytoplasm			5
ppnK	Probable inorganic polyphosphate/ATP-NAD kinase	2.7.1.23	P	E	No		2 / 0	Inhibitors for *Ec* and others (BRENDA)	295	Cytoplasm	Spectrophotometric assay (Mori *et al.*, 2005)	7 structures, dimer, 4-mer, apo and ligand-bound, *Yp* (2.0 - 3.0 Å)	1
priA	Primosome assembly protein PriA	3.6.1.15?	E	E	No		0 / 0	Inhibitors for *N. gonorrhoeae* (Sunchu *et al.*, 2012)	739	Cytoplasm	ATPase/dATPase: Radioactivity (Sunchu *et al.*, 2012). ATP-dependent helicase: DNA binding assay (Sunchu et al., 2012)	2 structures, monomer, apo and ligand-bound (2.7 - 4.0 Å)	5
proS	Prolyl-tRNA synthetase	6.1.1.15	E	E	Yes, 48%	0.57	4 / 0	Inhibitors for *Ec* (nM) (BRENDA)	571	Cytoplasm	Radioactivity (Stathopoulos *et al.*, 2001)	11 structures, monomer, apo, ligand-bound, complexed with other proteins (1.5 - 3.0 Å)	4
pyrG	CTP synthase	6.3.4.2	E	E	Yes, 45%		0 / 0	Inhibitors for *Ec* (mM and nM) (BRENDA)	542	Cytoplasm	Spectrophotometric assay (Lunn and Bearne, 2004)	8 structures, dimer, 4-mer, apo and ligand-bound (1.5 - 3.5 Å)	5
pyrH	Uridylate kinase	2.7.4.22	E	E	No		0 / 0	Inhibitors for *Ec* (mM and nM): ATP, GTP, UMP, UTP (BRENDA)	245	Cytoplasm	Spectrophotometric coupled assay. (Bucurenci *et al.*, 1998)	11 structures, dimer, apo and ligand-bound, *Ba* (1.5 - 3.5 Å)	1
serS	Seryl-tRNA synthetase	6.1.1.11	E	E	Yes, 33%		1 / 0	Inhibitors (mM) (BRENDA)	426	Cytoplasm	Radioactivity (Landeka *et al.*, 2000)	5 structures, dimer, apo and ligand-bound (2.0 - 3.0 Å)	5
spoT	Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase	2.7.6.5 3.1.7.2	E	E	No		2 / 0	None	701	Cytoplasm	GTP pyrophosphokinase: Radioactivity (Avarbock *et al.*, 2005). Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase: Radioactivity (Avarbock *et al.*, 2005)	5 structures, dimer, monomer, apo and ligand-bound (1.5 - 2.5 Å)	5
sucA	2-oxoglutarate dehydrogenase E1	1.2.4.2	E	E	Yes, 42%		3 / 0	None	943	Cytoplasm	Spectrophotometric assay (Kresze and Ronft, 1981)	12 structures, dimer, apo, ligand-bound, complexed with other proteins (1.5 - 3.0 Å)	5
sucB	Dihydrolipoamide succinyltransferase	2.3.1.61	E	E	Yes, 58%		0 / 0	Inhibitor for *Hs* and *Rn*: Coenzyme Q0 (BRENDA)	409	Cytoplasm	Spectrophotometric coupled assay (Hirashima *et al.*, 1967; Koike *et al.*, 2000)	None	5
thiL	Thiamine monophosphate kinase	2.7.4.16	E	E	No		0 / 0	Inhibitors for *Ec* (mM and unknown K_i) (BRENDA)	322	Cytoplasm	Manometric assay (CO_2 liberation from pyruvate) (Nishino, 1972). Radioactivity (Iwashima *et al.*, 1986)	6 structures, dimer, apo and ligand-bound (1.5 - 3.0 Å)	5

Yp gene name	Gene product	EC	Essentiality		Human homolgue	Druggability			Tractability as a target				
			Pa	DEG		QED	ChEMBL	Drug-like compounds	Size (AA)	Location	Assay	Structures (resolution)	XtalPred
thrS	Threonyl-tRNA synthetase	6.1.1.3	P	E	Yes, 39%	0.24	1 / 0	Inhibitors for *Ec* (nM, Ruan *et al.*, 2005) and *Sc* (no K_i values, BRENDA)	640	Cytoplasm	Fluorescence assay (Bovee *et al.*, 2003), Radioactivity (Ruan *et al.*, 2005)	18 structures, dimer, apo and ligand-bound (1.5 - 3.5 Å)	3
thyA	Thymidylate synthase	2.1.1.45	E	E	Yes, 51%	0.76	14 / 0	Inhibitors for different *sp.* (nM) (BRENDA)	264	Cytoplasm	Spectrophotometric assay (Haertle *et al.*, 1979; Kawate *et al.*, 2002)	127 structures, dimer, apo and ligand-bound (1.3 - 3.0 Å)	1
tmk	Thymidylate kinase	2.7.4.9	E	E	No	0.31	1 / 0	Inhibitors for different *sp.* (nM and mM) (BRENDA)	210	Cytoplasm	*Y. pestis*: Spectrophotometric coupled assay. (Chenal-Francisque *et al.*, 1999)	50 structures, monomer, apo and ligand-bound (1.3 - 3.0 Å)	3
trmD	tRNA (guanine-N(1)-)-methyltransferase	2.1.1.228	E	E	No		0 / 0	Inhibitors for different *sp.* (nM) (BRENDA)	252	Cytoplasm	Radioactivity (Takeda *et al.*, 2006)	22 structures, dimer, apo and ligand-bound (1.3 - 3.0 Å)	3
ubiB	ubiquinone biosynthesis protein	2.7.-.-	E	E	Yes, 30%		2 / 0	None	533	Inner membrane	Function of UbiB remains unclear but there is some evidence that suggests it might act as a kinase with ATPase activity (Aussel *et al.*, 2013)	3 structures, 4-mer, apo (2.0 - 3.0 Å)	5
ubiG	3-demethylubiquinone-9 3-methyltransferase	2.1.1.222 2.1.1.64	E	E	Yes, 35%	0.55	2 / 0	None	232	Cytoplasm	Assays for 5-demethylubiquinone-9-methyltransferase from rat liver mitochondria; 2-polyprenyl-6-hydroxyphenol methylase: Calorimetric assay. 3-demethylubiquinone 3-O-methyltransferase: Radioactivity (Houser and Olson, 1977).	2 structures, monomer, apo and complexed with other protein (2.0 Å)	1
uppS	Undecaprenyl pyrophosphate synthase	2.5.1.31	E	E	Yes, 36%	0.64	1 / 0	Inhibitors for *Ec* and *Ml*(nM), and others (no K_i values) (BRENDA)	251	Cytoplasm	Radioactivity (Chen *et al.*, 2005)	33 structures, monomer, apo and ligand-bound (1.3 - 3.0 Å)	1

APPENDIX B

Structural comparisons of KynB monomers and dimers.

Table B.1: Results of the C_α overlay between *Ba*KyB and *Bc*KynB subunits. The number of atoms used are: 206 for *Ba*KynB (subunits A and B), 207 for *Ba*KynB (subunit C and D), and 207 for *Bc*KynB subunits.

		*Ba*KynB				*Bc*KynB			
		A	B	C	D	A	B	C	D
*Ba*KynB	A	-	0.387	0.130	0.370	1.172	1.195	1.223	1.217
	B		-	0.348	0.100	1.195	1.252	1.222	1.278
	C			-	0.343	1.206	1.234	1.193	1.258
	D				-	1.165	1.221	1.258	1.246
*Bc*KynB	A					-	0.269	0.178	0.301
	B						-	0.210	0.129
	C							-	0.255
	D								-

Table B.2: Results of the C_α overlay between *Pa*KynB, *Ba*KyB and *Bc*KynB subunits. The number of atoms used are: 206 for *Pa*KynB (*Ba*KynB and *Bc*KynB as indicated in Table B.1).

		*Pa*KynB	
		A	B
*Ba*KynB	A	1.183	1.178
	B	1.212	1.172
	C	1.203	1.210
	D	1.153	1.167
*Bc*KynB	A	0.694	0.622
	B	0.704	0.630
	C	0.684	0.605
	D	0.739	0.664
*Pa*KynB	A	-	0.324
	B		-

Table B.3: Results of the C_α overlay between *Pa*KynB, *Ba*KyB and *Bc*KynB dimers. The number of atoms used are indicated in Tables B.1-2.

		*Ba*KynB		*Bc*KynB		*Pa*KynB
		A/B	C/D	A/B	C/D	A/B
*Ba*KynB	A/B	-	0.156	1.255	1.217	1.195
	C/D		-	1.254	1.226	1.204
*Bc*KynB	A/B			-	0.160	0.698
	C/D				-	0.714
*Pa*KynB	A/B					-

APPENDIX C

Resume of the work carried out with other potential targets; FabZ, FtsZ, MsbA, RsmH, PyrG and NorM.

C.1 Introduction

Additional targets were also included in the program. Only limited amount of work was carried out on them due to difficulties of expression, purification or crystallisation, and also the prioritisation of other proteins.

C.1.1 FabZ

Fatty acids are essential for membrane formation and its biophysical properties, which vary depending on the amount of saturated and unsaturated compounds. The fatty acid biosynthetic pathway type I (FASI) is present in mammals and it is carried out by one gene encoding for a multifunctional polypeptide. In bacteria, plants and parasites, the fatty acid synthesis type II (FASII) independent genes encode for mono-functional proteins (White *et al.,* 2005). Due to these differences, proteins involved in the FASII have been proposed as selective potential antibacterial drug targets and inhibitors have been characterised for various FASII enzymes (Hiltunen *et al.,* 2008; Yao and Rock, 2015). FabZ is the 3-(R)-hydroxyacyl-[Acyl-Carrier-Protein] dehydratase from the FASII. It catalyses the dehydration of 3(R)-hydroxyacyl-ACP to trans-2-enoyl-ACP (Heath and Rock, 1996; White *et al.,* 2005). The quaternary structure of FabZ is a trimer of dimers and inhibitors have been described for *H. pylori*, *Plasmodium falciparum* and *P. aeruginosa* (Tasdemir *et al.,* 2006; Moynié *et al.,* 2013).

C.1.2 FtsZ

The tubulin homologue FtsZ is essential for cell division in bacteria (Errington *et al.,* 2003). It presents a similar secondary structure to tubulin and GTPase activity occurs at the N-terminal region. FtsZ, as well as tubulin, forms protofilaments with the GTP bound protein form. Differences are noted between these proteins; while tubulin has two distinct subunits (α and β) that associate longitudinally to form the filaments, there is only one FtsZ subunit that uses lateral association. This together with a sequence identity of only 20%, indicate selectivity towards the bacterial protein may be feasible (Ma and Ma, 2012). The FtsZ filament forms the Z-ring, an essential structure for cytokinesis in bacteria, at the equatorial plane of the cell that serves for the recruitment of at least 12 cell-division proteins to form what is termed the divisome. The GTPase activity of FtsZ is formed at the interface of two adjacent subunits and the monomers dissociate from the filaments by GTP hydrolysis to GDP (Oliva and Löwe, 2007; Ma and Ma, 2012). Good inhibitors have been characterised for FtsZ but none of them are optimised for clinical use (Ma and Ma, 2012; Blaauwen *et al.,* 2014).

C.1.3 MsbA

Lipopolysaccharides are constituted of an *O*-polysaccharide and the lipid A core. The dimeric membrane protein MsbA is the lipid A export ATP-binding permease and belongs to the ABC transporter family. It mediates the translocation (or flipping) of lipid A-core from the cytoplasmic side of the inner membrane to the periplasmic space (Ghanei *et al.,* 2007). MsbA lipid translocation is ATP-hydrolysis dependent, and the mechanism is still unclear (Kaul and Pattan, 2011). *P. aeruginosa* MsbA has been subject of study in the Hunter lab before and it was included in this work to gain experience in expression, purification and crystallisation of membrane proteins and to

assess its incorporation as a target for the biowarfare bacteria of interest.

C.1.4 RsmH

RsmH is the AdoMet-dependent 5-methyluridine methyltransferase. It catalyses the *N*4-methylation of the ribosomal RNA (rRNA) Cys1402. Post-translational modifications of rRNA play important roles in protein synthesis quality control (Wei *et al.,* 2012) but it is still unclear if it is essential or not (Kyuma et al., 2015). Preliminary data from Dstl (see Part Two) suggested RsmH is essential for the bacteria of interest. Additionally, inhibitors targeting RsmH have not been developed so it was included for characterisation of new antibacterial drug targets in *Y. pestis*, *B. pseudomallei* and *B. anthracis*.

C.1.5 PyrG

CTP synthase, also called PyrG, is the last enzyme involved in *de novo* CTP biosynthesis. Some bacteria and parasites present only one route for *de novo* CTP synthesis while in mammals there exists an additional salvage pathway (Huang and Graves, 2002). PyrG is an ammonia-dependent aminotransferase containing two differentiated domains. The N-terminal domain comprises the CTP synthase function and the C-terminal the glutaminase activity (Lunn and Bearne, 2004; Goto *et al.,* 2004). The glutaminase domain hydrolyses L-glutamine to generate ammonia. This is transferred through a tunnel connecting both active sites to the N-terminal domain and the ammonia is transferred to the phosphorylated-UTP (Endrizzi *et al.,* 2005). The reaction can also occur in the absence of L-glutamine but in the presence of free ammonia. PyrG activity is highly regulated. On one hand, ATP and CTP are allosteric activators of the CTP synthase activity while GTP induces the binding of L-glutamate.

On the other hand, GTP also inhibits PyrG when its concentration is greater than 0.15 mM in *E. coli*. Additionally, product inhibition by CTP is also observed (Lunn and Bearne, 2004). PyrG can be present as monomer, dimer and tetramer depending on the protein concentration (Goto *et al.,* 2004). Dimers and tetramers can also form fibrilar structures, called micro-cytoophidia, which can associate and generate macro-cytoophidia structures (Liu, 2011). These macrostructures exist as a way of regulating PyrG activity as they are disassembled under high CTP demand situations (Noree *et al.,* 2014). The active form of PyrG is a tetramer and is formed when ATP and UTP bind to a dimer. Inhibitors of PyrG have been described for various species (Hofer *et al.,* 2001; Endrizzi *et al.,* 2005; Yoshida *et al.,* 2012; Mori *et al.,* 2015).

C.1.6 NorM

The last selected target is the membrane transporter NorM or PmpM. It is a multidrug resistance protein similar to MATE (multidrug and toxic compounds extrusion) efflux pump family (Su *et al.,* 2008). Efflux pumps are known to be involved in drug resistance for various bacteria including *B. pseudomallei* (Schweizer, 2003; Biot *et al.,* 2011). This protein has been selected as a new strategy for a potential drug target for *B. pseudomallei* (*Bp*NorM). *Bp*NorM might be essential under the chemical stress conditions caused by drugs, so its inhibition might play an important role to avoid drug resistance.

C.2 Experimental procedures

Experimental details are shown in Table C.1.

Genes encoding for the proteins in Table C.1 were cloned into a pET15bTEV vector using the restriction enzymes NdeI (at the 3’ DNA end) and BamHI (5’ end), and

sequenced to check gene integrity. Expression and purification protocols were similar to those detailed previously; 1) First Ni-NTA column, 2) His_6-tag cleavage and 3) second Ni-NTA column. Cell lysis buffers and buffers A for Ni-NTA columns are detailed in Table C.1, and buffer B consisted of buffer A plus 0.8 M imidazole (0.3 M imidazole in the case of MsbA). To cleave His_6-tags, 1 mg of TEV protease was added per 10 mg of protein and incubated at 4ºC O/N. Protein concentration was determined using the predicted molar extinction coefficient (ε, PROTPARAM, Gasteiger *et al.,* 2003) and the Lambert-Beer equation.

*Pa*MsbA expression and purification protocols had been optimised previously by Dr. A. Dawson. Cell cultures were firstly centrifuged at 4,000 *g* for 40 minutes. Then pellets were resuspended in the lysis buffer. Cell lysis was carried out using the French Press, insoluble debris removed by centrifugation at 12,000 *g*, and then membrane isolation was obtained after centrifugation for 1 hour at 208,000 *g*. The pellet was then solubilised in buffer A, containing 0.02% *n*-dodecyl β-D-maltoside (DDM) detergent, and purification was pursued as indicated above.

Fragile crystals were obtained for *Yp*PyrG from condition D1 from The Classics Suite. Optimisation plates were prepared using different protein concentrations, temperature (4ºC, 20ºC and room temperature), including different concentrations of cryo-protectants in the drops (ethylene glycol, PEG 400, glycerol and MPD), and different precipitants (tri-sodium citrate, tri-potassium citrate and ammonium sulfate) were also tested based on the commercial screening results. Severe aggregation or no crystals were observed in all cases. Hanging and sitting drop plates, as well as microbatch silicon oil plates were used. While mounting crystals, different cryo-protectants were

tested, including ethylene glycol, glucose, MPD, PEG 400 and glycerol but crystals dissolved or cracked. To minimize crystal manipulation, counter diffusion using 0.1 mm capillaries was carried out. Optimisation of a different condition (H11 from the PEGs Suite) was started and less fragile crystals obtained, however they were too small to be tested. Renata Reis, a visiting PhD student from the University of Sao Paulo, is carrying out further optimisation for the H11 condition. Improvements in the crystal size and diffraction to 9 Å has been obtained so far and further optimisation of this condition will follow. Ligands were also included for co-crystallisation. Different combinations of 10 mM ATP, UTP, CTP and L-glutamine were tested. DSF was used to check protein stability of *Yp*PyrG with and without His_6-tag after a freeze-thaw process, and fresh *Yp*PyrG (without His_6-tag). DSF was carried out with 5 μM of protein in each well, in triplicate. SYPRO orange was added to the samples in buffer A. The temperature range tested goes from 25ºC to 94º, and data were analysed using the Boltzmann sigmoidal equation with GrapPad PRISM (www.graphpad.com) to determine the T_m values. Samples were also checked using SDS-PAGE. A DDU fragment library screening was carried out with *Yp*PyrG enzyme.

Protein	Organism (MW, ε)	UniProt ID	Gene size (bp)	Expression	Lysis buffer, Ni-NTA (buffer A)	Crystallisation	Quaternary structure	Fragment screening
FabZ^{GS}	*Ba* (16.1, 3105)	Q81JE0	441	Various *E. coli* strains, temperatures and IPTG concentrations *[1]	-	-	-	-
FtsZ^{GS}	*Yp* (40.4, 4470)	Q7CGB3	1177	BL21(DE3), 1 mM IPTG, 20ºC O/N	Various buffers	-	-	-
	Bp (41.6, 13075)	Q63QK2	1202	BL21(DE3), 1 mM IPTG, 20ºC O/N	Various buffers	-	-	-
	Ba (40.9, 1490)	Q81WD4	1141	BL21(DE3), 1 mM IPTG, 20ºC O/N	Various buffers	-	-	-
MsbA^{GS}	*Pa* (66.4, 58330)	Q9HUG8	1809	C43(DE3) 1 mM IPTG, room temperature O/N	50 mM Na_2HPO_4 pH 7.9, 300 mM NaCl, 2 mM MgCl2, 0.02% DDM, 10% Glycerol	Lipidic sponge phase screen	SEC-MALS *[3]	-
RsmH^{GW}	*Yp* (35.6, 16960)	Q8ZIF7	969	BL21(DE3), 1 mM IPTG, 20ºC O/N	20 mM Tris pH 7.4, 200 mM NaCl	Various screens	-	-
	Bp (34.2, 9970)	Q63QI9	948	BL21(DE3), 1 mM IPTG, 20ºC O/N	50 mM Tris pH 7.4, 500 mM NaCl (± 10% (v/v) glycerol)	-	-	-
	Ba (35.0, 19035)	Q81WC3	939	Various *E. coli* strains, temperatures and IPTG concentrations *2	-	-	-	-
PyrG^{GW}	*Yp* (60.4, 45310)	Q8ZBN1	1675	BL21(DE3), 1 mM IPTG, 20ºC O/N	20 mM Tris pH 7.4, 200 mM NaCl	Various techniques	SEC *[4]	20 µM of His_6-tagged protein, ATP and TEV as controls.
	Ba (59.7, 48040)	Q81JW1	1614	BL21(DE3), 1 mM IPTG, 20ºC O/N	20 mM Tris pH 7.4, 200 mM NaCl	Various screens	SEC *[4]	-
NorM^{GW}	*Bp* (48.2, 66920)	Q63S57	1432	C43(DE3) 1 mM IPTG, 20ºC O/N	Purification not carried out	-	-	-

Table C.1: Material procedures information for FabZ, FtsZ, MsbA, RsmH, PyrG and NorM from *B. anthracis* (*Ba*), *Y. pestis* (*Yp*), *B. pseudomallei* (*Bp*), *P. aeruginosa* (*Pa*). (^{GS}) Encoding genes purchased from GeneStript. (^{GW}) Encoding genes purchased from GeneWizt. Protein molecular weight (MW) is shown in kDa and the ε in $M^{-1}cm^{-1}$. (*[1]) *E. coli* BL21(DE3), BL21(DE3)pLysS, C43(DE3), Arctic Express (DE3), Rosetta (DE3), 15ºC, 20ºC, room temperature, IPTG concentrations form 0.4 to 1 mM, and from 6 hours to O/N expression. (*[2]) *E. coli* BL21(DE3), BL21(DE3)pLysS, 15ºC, 20ºC, room temperature, IPTG concentrations form 0.4 to 1 mM, O/N expression. (*[3]) SEC with Multi-Angle Light Scattering using column Superdex 200 HiLoad 16/60 pg. (*[4]) Using column Superdex 200 HiLoad 16/60 pg, and native-PAGE 3-12% (Invitrogen gels). (-) Not applicable.

C.3 Results

Results are summarised in Table C.2.

Protein	Organism	Expression, protein yield (mg L^{-1})	Crystallisation	Quaternary structure	Fragment screening
FabZ	*Ba*	Not soluble, possibly toxic	-	-	-
FtsZ	*Yp*	Aggregation	-	-	-
	Bp	Aggregation	-	-	-
	Ba	Aggregation	-	-	-
MsbA	*Pa*	3	Poor diffracting crystals	Dimer	-
RsmH	*Yp*	10	No crystals	-	-
	Bp	Aggregation	-	-	-
	Ba	Not soluble	-	-	-
PyrG	*Yp*	30	Fragile crystals	Dimer	32 hits
	Ba	5	Small needles	Dimer	-
NorM	*Bp*	C43(DE3)	-	-	-

Table C.2: Results resume from work carried out on FabZ, FtsZ, MsbA, RsmH, PyrG and NorM from *B. anthracis* (*Ba*), *Y. pestis* (*Yp*), *B. pseudomallei* (*Bp*) and *P. aeruginosa* (*Pa*). (-) Not carried out.

In the case of *Yp*PyrG, variable crystal forms were observed depending on protein concentration. This may be due to the quaternary structure state in the protein solution as it has been described that formation of dimers and tetramers is concentration-dependent (Goto *et al.,* 2004). Crystals either too fragile or small (less than 10 μm) were obtained. During crystal optimisation, it was noticed that crystals grew solely from fresh protein. The SDS-PAGE for frozen-thawed protein (f-*Yp*PyrG) and fresh samples (*Yp*PyrG) did not show degradation or differences. Additionally, DSF results indicated the same stability with T_m of 66.5ºC (f-*Yp*PyrG) and 64.3ºC (*Yp*PyrG). These results suggested that the freezing process interferes in the crystallisation process by a mechanism that is not understood. In the case of *Ba*PyrG, optimisation was carried out for different conditions but no improvement was made. No further optimisation was attempted. From the *Yp*PyrG fragment screen, 32 potential hits were found but further validation has not being carried out. Hits are shown in Table C.3.

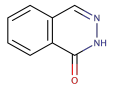
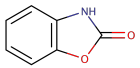
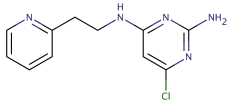
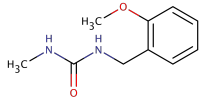
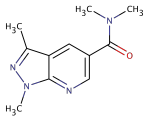
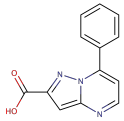
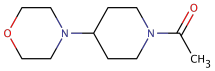
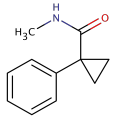
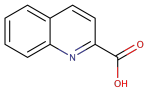
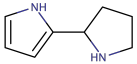
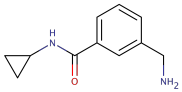
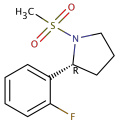
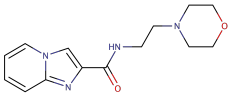
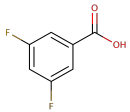
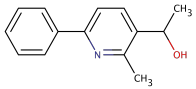
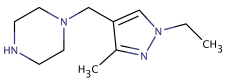
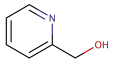
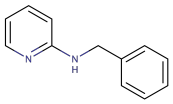
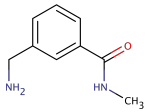
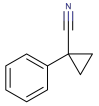
#	Compound/ Structure	#	Compound/ Structure	#	Compound/ Structure	#	Compound/ Structure
1	1(2H)-phthalazinone 	9	2-benzoxazolinone 	17	6-chloro-*N*'-[2-(2-pyridyl)ethyl]pyrimidine-2,4-diamine 	25	3-[(2-methoxyphenyl)methyl]-1-methyl-urea 
2	*N*,*N*,1,3-tetramethylpyrazolo[3,4-b]pyridine-5-carboxamide 	10	7-phenylpyrazolo[1,5-a]pyrimidine-2-carboxylic acid 	18	1-(4-morpholin-4-ylpiperidin-1-yl)ethanone 	26	*N*-methyl-1-phenyl-cyclopropanecarboxamide 
3	quinoline-2-carboxylic acid 	11	2-pyrrolidin-2-yl-1H-pyrrole 	19	3-(aminomethyl)-*N*-cyclopropylbenzamide hydrochloride 	27	(2R)-2-(2-fluorophenyl)-1-methylsulfonyl-pyrrolidine 
4	*N*-(2-morpholinoethyl) imidazo[1,2-a]pyridine-2-carboxamide 	12	3,5-difluorobenzoic acid 	20	2-methyl-6-phenylnicotinic acid11 	28	1-(1-ethyl-3-methyl-1 H -pyrazol-4-ylmethyl)-piperazine 
5	2-pyridinemethanol 	13	2-benzylaminopyridine 	21	3-(aminomethyl)-*N*-methylbenzamide hydrochloride 	29	1-(pyridin-2-yl)cyclopropanecarbonitrile 

Table C.3: Hits from *Yp*PyrG fragment screening.

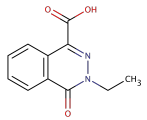
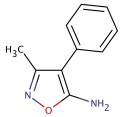
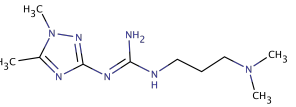
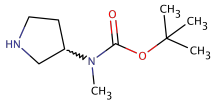
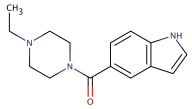
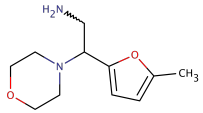
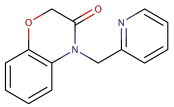
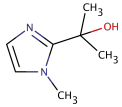
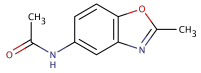
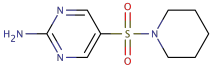
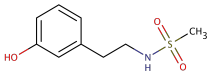
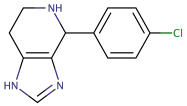
#	Compound/ Structure	#	Compound/ Structure	#	Compound/ Structure	#	Compound/ Structure
6	3-ethyl-4-oxo-3,4-dihydro-phthalazine-1-carboxylic acid 	14	3-methyl-4-phenylisoxazol-5-amine 	22	3-[2-(1,5-dimethyl-1,2,4-triazol)guanidio]-1-*N*-dimethyl-propane 	30	3-(*N*-tert-butoxycarbonyl-*N*-methylamino)pyrrolidine 
7	Methanone, (4-ethyl-1-piperazinyl)-1H-indol-5-yl- 	15	2-(5-methyl-furan-2-yl)-2-morpholin-4-yl-ethylamine 	23	6-amino-4-(2-pyridylmethyl)-1,4-benzoxazin-3-one 	31	2-(1-methyl-1H-imidazol-2-yl)propan-2-ol 
8	*N*-(2-methyl-1,3-benzoxazol-5-yl)acetamide 	16	5-(1-piperidylsulfonyl)pyrimidin-2-amine 	24	*N*-[2-(3-methoxyphenyl)ethyl]methanesulfonamide 	32	4-(4-chlorophenyl)-4,5,6,7-tetrahydro-3H-imidazo[4,5-c]pyridine 

Table C.3: Continuation.

REFERENCES

Aherne, W., Maloney, A., Prodromou, C., Rowlands, M.G., Hardcastle, A., Boxall, K., Clarke, P., Walton, M.I., Pearl, L. and Workman, P. (2003) Assays for HSP90 and inhibitors. *Methods in Molecular Medicine*, 85, 149-161.

Alanis, A.J. (2005) Resistance to antibiotics: are we in the post-antibiotic era? *Archives of Medical Research*, 36, 697-705.

Alexopoulos, J.A., Guarné, A. and Ortega, J. (2012) ClpP: a structurally dynamic protease regulated by AAA+ proteins. *Journal of Structural Biology*, 179, 202-210.

Allen, F.H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B*, 58, 380-388.

Arndt, J.W., Schwarzenbacher, R., Page, R., Abdubek, P., Ambing, E., Biorac, T., Canaves, J.M., Chiu, H.J., Dai, X., Deacon, A.M., DiDonato, M., Elsliger, M.A., Godzik, A., Grittini, C., Grzechnik, S.K., Hale, J., Hampton, E., Han, G.W., Haugen, J., Hornsby, M., Klock, H.E., Koesema, E., Kreusch, A., Kuhn, P., Jaroszewski, L., Lesley, S.A., Levin, I., McMullan, D., McPhillips, T.M., Miller, M.D., Morse, A., Moy, K., Nigoghossian, E., Ouyang, J., Peti, W.S., Quijano, K., Reyes, R., Sims, E., Spraggon, G., Stevens, R.C., van den Bedem, H., Velasquez, J., Vincent, J., von Delft, F., Wang, X., West, B., White, A., Wolf, G., Xu, Q., Zagnitko, O., Hodgson, K.O., Wooley, J. and Wilson, I.A. (2005) Crystal structure of an alpha/beta serine hydrolase (YDR428C) from *Saccharomyces cerevisiae* at 1.85 Å resolution. *Proteins: Structure, Function and Bioinformatics*, 58, 775-778.

Anderson, P.D. and Bokor, G. (2012) Bioterrorism: pathogens as weapons. *Journal of Pharmacy Practice*, 25, 521-529.

Anderson, M.S., Bull, H.G., Galloway, S.M., Kelly, T.M., Mohan, S., Radika, K. and Raetz, C.R. (1993) UDP-N-acetylglucosamine acyltransferase of *Escherichia coli*. The first step of endotoxin biosynthesis is thermodynamically unfavorable. *Journal of Biological Chemistry*, 268, 19858-19865.

Anisimov, A.P. and Amoako, K.K. (2006) Treatment of plague: promising alternatives to antibiotics. *Journal of Medical Microbiology*, 55, 1461-1475.

Ataide, S.F. and Ibba, M. (2004) Discrimination of cognate and noncognate substrates at the active site of class II lysyl-tRNA synthetase. *Biochemistry*, 43, 11836-11841.

Aussel, L., Loiseau, L., Hajj Chehade, M., Pocachard, B., Fontecave, M., Pierrel, F. and Barras, F. (2014) ubiJ, a new gene required for aerobic growth and proliferation in macrophage, is involved in coenzyme Q biosynthesis in *Escherichia coli* and *Salmonella enterica* serovar t*yphimurium*. *Journal of Bacteriology*, 196, 70-79.

Avarbock, A., Avarbock, D., The, J.S., Buckstein, M., Wang, Z.M. and Rubin, H. (2005) Functional regulation of the opposing (p)ppGpp synthetase/hydrolase activities of RelMtb from *Mycobacterium tuberculosis*. *Biochemistry*, 44, 9913-9923.

Baell, J.B., Holloway, G.A. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry*, 53, 2719-2740.

Baran, H. and Kepplinger, B. (2014) D-cycloserine lowers kynurenic acid formation--new mechanism of action. *European Neuropsychopharmacology: the Journal of the European College of Neuropsychopharmacology*, 24, 639-644.

Barnes, M.J., Estlin, E.J., Taylor, G.A., Aherne, G.W., Hardcastle, A., McGuire, J.J., Calvete, J.A., Lunec, J., Pearson, A.D. and Newell, D.R. (1999) Impact of polyglutamation on sensitivity to raltitrexed and methotrexate in relation to drug-induced inhibition of *de novo* thymidylate and purine biosynthesis in CCRF-CEM cell lines. *Clinical Cancer Research*, 5, 2548-2558.

Barras, V. and Greub, G. (2014) History of biological warfare and bioterrorism. *Clinical Microbiology and Infection*, 20, 497-502.

Bartling, C.M. and Raetz, C.R. (2008) Steady-state kinetics and mechanism of LpxD, the N-acyltransferase of lipid A biosynthesis. *Biochemistry*, 47, 5290-5320.

Baugh, L., Gallagher, L.A, Patrapuvich, R., Clifton, M.C., Gardberg, A.S., Edwards, T.E., Armour, B., Begley, D.W., Dieterich, S.H., Dranow, D.M., Abendroth, J., Fairman, J.W., Fox, D.III, Staker, B.L., Phan, I., Gillespie, A., Choi, R., Nakazawa-Hewitt, S., Nguyen, M.T., Napuli, A., Barrett, L., Buchko, G.W., Stacy, R., Myler, P.J., Stewart, L.J., Manoil, C., van Voorhis, W.C. (2013) Combining functional and structural genomics to sample the essential *Burkholderia* structome. *PLoS One*, 8, online.

Baugh, L., Phan I., Begley, D.W., Clifton, M.C., Armour, B., Dranow, D.M, Taylor, B.M., Muruthi, M.M., Abendroth, J., Fairman, J.W., Fox III, D., Dieterich S.H., Staker, B.L., Gardberg, A.S., Choi, R., Hewitt, S.N., Napuli, A.J., Myers, J., Barrett, L.K., Zhang, Y., Ferrell, M., Mundt, E., Thompkins, K., Tran, N., Lyons-Abbott, S., Abramov, A., Sekar, A., Serbzhinskiy, D., Lorimer, D., Buchko, G.W., Stacy, R., Stewart, L.J., Edwards, T.E., van Voorhis, W.C. and Myler, P.J.

(2015) Increasing the structural coverage of tuberculosis drug targets. *Tuberculosis*, 95, 142-148.

Bebrone, C. (2007) Metallo-β-lactamases and their superfamily (classification, activity, genetic organization, structure, zinc coordination). *Biochemical Pharmacology*, 74, 1686–1701.

Bento, A.P., Gaulton, A., Hersey, A., Bellis, L.J., Chambers, J., Davies, M., Krüger, F.A., Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R. and Overington, J.P. (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, 42, 1083-90.

Bermingham, A. and Derrick, J.P. (2002) The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. *Bioessays*, 24, 637-648.

Bernier, S., Dubois, D.Y., Therrien, M., Lapointe, J. and Chênevert. R. (2000) Synthesis of glutaminyl adenylate analogues that are inhibitors of glutaminyl-tRNA synthetase. *Bioorganic and Medicinal Chemistry Letters*, 10, 2441-2444.

Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S. and Hopkins, A.L. (2012) Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4, 90-98.

Bienvenue, D.L., Gilner, D.M., Davis, R.S., Bennett, B. and Holz, R.C. (2003) Substrate specificity, metal binding properties, and spectroscopic characterization of the DapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase from *Haemophilus influenzae*. *Biochemistry*, 42, 10756-10763.

Bigley, A.N. and Raushel, F.M. (2013) Catalytic mechanisms for phosphotriesterases. *Biochimica et Biophysica Acta*, 1834, 443-453.

Biot, F.V., Valade, E., Garnotel, E., Chevalier, J., Villard, C., Thibault, F.M., Vidal, D.R. and Pagès, J.M. (2011) Involvement of the efflux pumps in chloramphenicol selected strains of *Burkholderia thailandensis*: proteomic and mechanistic evidence. *PLoS One*, 6, online.

Bjerregaard-Andersen, K., Sommer, T., Jensen, J., Jochimsen, B., Etzerodt, M. and Morth, J.P. (2014) A proton wire and water channel revealed in the crystal structure of isatin hydrolase. *Journal of Biological Chemistry*, 289, 21351-21359.

Blaauwen, T., Andreu, J.M. and Monasterio, O. (2014) Bacterial cell division proteins as antibiotic targets. *Bioorganic Chemistry*, 55, 27-38.

Bognar, A.L., Osborne, C., Shane, B., Singer, S.C. and Ferone, R. (1985) Folylpoly-gamma-glutamate synthetase-dihydrofolate synthetase. Cloning and high expression of the *Escherichia coli* folC gene and purification and properties of the gene product. *Journal of Biological Chemistry*, 260, 5625-5630.

Bond, C.S. and Schüttelkopf, A.W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallographica Section D*, 65, 510-512.

Bosch, F. and Rosich, L. (2008) The contributions of Paul Ehrlich to pharmacology: a tribute on the occasion of the centenary of his Nobel Prize. *Pharmacology*, 82, 171-179.

Böttcher, T. and Sieber, S.A. (2008) Beta-lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of *Staphylococcus aureus*. *Journal of American Chemical Society*, 130, 14400-14401.

Böttcher, T. and Sieber, S.A. (2009) Beta-lactones decrease the intracellular virulence of *Listeria monocytogenes* in macrophages. *ChemMedChem*, 4, 1260-1263.

Bougie, J. M. (2011) Expression, purification, and characterization of kynurenine formamidase from *Bacillus cereus*. Masters Thesis, George Manson University, Fairfax, VA, U.S.A.

Bouhss, A., Dementin, S., Parquet, C., Mengin-Lecreulx, D., Bertrand, J.A., Le Beller, D., Dideberg, O., van Heijenoort, J. and Blanot, D. (1999) Role of the ortholog and paralog amino acid invariants in the active site of the UDP-MurNAc-L-alanine:D-glutamate ligase (MurD). *Biochemistry*, 28, 12240-12247.

Bouhss, A., Trunkfield, A.E., Bugg, T.D. and Mengin-Lecreulx, D. (2008) The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiology Reviews*, 32, 208-233.

Bovee, M.L., Pierce, M.A. and Francklyn, C.S. (2003) Induced fit and kinetic mechanism of adenylation catalyzed by *Escherichia coli* threonyl-tRNA synthetase. *Biochemistry*, 42, 15102-15113.

Brandish, P.E., Kimura, K.I., Inukai, M., Southgate, R., Lonsdale, J.T. and Bugg, T.D. (1996) Modes of action of tunicamycin, liposidomycin B, and mureidomycin A: inhibition of phospho-N-acetylmuramyl-pentapeptide translocase from *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, 40, 1640-1644.

Brooks, B.D. and Brooks, A.E. (2014) Therapeutic strategies to combat antibiotic resistance. *Advanced Drug Delivery Reviews*, 78, 14-27.

Brötz-Oesterhelt, H. and Sass, P. (2014) Bacterial caseinolytic proteases as novel targets for antibacterial treatment. *International Journal of Medical Microbiology*, 304, 23-30.

Bruning, J.B., Murillo, A.C., Chacon, O., Barletta, R.G. and Sacchettini, J.C. (2011) Structure of the *Mycobacterium tuberculosis* D-alanine:D-alanine ligase, a target

of the antituberculosis drug D-cycloserine. *Antimacrobial agents chemotherapy*, 55, 291-301.

Bucurenci, N., Serina, L., Zaharia, C., Landais, S., Danchin, A. and Bârzu, O. (1998) Mutational analysis of UMP kinase from *Escherichia coli*. *Journal of Bacteriology*, 180, 473-477.

Bugg, T.D.H., Braddick, D., Dowson, C.G. and Roper, D.I. (2011) Bacterial cell wall assembly: still an attractive antibacterial target. *Trends in Biotechnology*, 29, 167-173.

Bulik, D.A., van Ophem, P., Manning, J.M., Shen, Z., Newburg, D.S. and Jarroll, E.L. (2000) UDP-N-acetylglucosamine pyrophosphorylase, a key enzyme in encysting *Giardia*, is allosterically regulated. *Journal of Biological Chemistry*, 275, 14722-14728.

Buurman, E.T., Andrews, B., Gao, N., Hu, J., Keating, T.A., Lahiri, S., Otterbein, L.R., Patten, A.D., Stokes, S.S. and Shapiro, A.B. (2011) In vitro validation of acetyltransferase activity of GlmU as an antibacterial target in *Haemophilus influenzae*. *Journal of Biological Chemistry*, 286, 40734-40742.

Carlos, J.L., Paetzel, M., Brubaker, G., Karla, A., Ashwell, C.M., Lively, M.O., Cao, G., Bullinger, P. and Dalbey, R.E. (2000) The role of the membrane-spanning domain of type I signal peptidases in substrate cleavage site selection. *Journal of Biological Chemistry*, 275, 38813-38822.

Carvalho, C.L., Lopes de Carvalho, I., Zé-Zé, L., Núncio, M.S. and Duarte, E.L. (2014) Tularaemia: a challenging zoonosis. *Comparative Immunology, Microbiology and Infectious Diseases*, 37, 85-96.

Cava, F., Lam, H., de Pedro, M.A. and Waldor, M.K. (2010) Emerging knowledge of regulatory roles of D-amino acids in bacteria. *Cellular and Molecular Life Sciences*, 68, 817-831.

Chaikuad, A., Tacconi, E.M., Zimmer, J., Liang, Y., Gray, N.S., Tarsounas, M. and Knapp, S. (2014) A unique inhibitor binding site in ERK1/2 is associated with slow binding kinetics. *Nature Chemical Biology*, 10, 853-860.

Chang, H.H., Cohen, T., Grad, Y.H., Hanage, W.P., O'Brien, T.F. and Lipsitch, M. (2015) Origin and proliferation of multiple-drug resistance in bacterial pathogens. *Microbiology and Molecular Biology Reviews*, 79, 101-116.

Chang, A., Schomburg, I., Placzek, S., Jeske, L., Ulbrich, M., Xiao, M., Sensen, C.W. and Schomburg, D. (2015) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Research*, 43, 439-46.

Chen, V.B., Arendall III, W.B., Headd, J.J., Keedy, D.A., Immormino, G.J.K., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D*, 66, 12-21.

Chen, A.P., Chang, S.Y., Lin, Y.C., Sun, Y.S., Chen, C.T., Wang, A.H. and Liang, P.H. (2005) Substrate and product specificities of cis-type undecaprenyl pyrophosphate synthase. *Biochemical Journal*, 386, 169-176.

Chen, L., Oughtred, R., Berman, H.M. and Westbrook, J. (2004) TargetDB: a target registration database for structural genomics projects. *Bioinformatics*, 20, 2860-2862.

Chenal-Francisque, V., Tourneux, L., Carniel, E., Christova, P., Li de la Sierra, I., Bârzu, O. and Gilles, A.M. (1999) The highly similar TMP kinases of *Yersinia*

pestis and *Escherichia coli* differ markedly in their AZTMP phosphorylating activity. *European Journal of Biochemistry*, 265, 112-119.

Childs, W.C. and Neuhaus, F.C. (1980) Biosynthesis of D-alanyl-lipoteichoic acid: characterization of ester-linked D-alanine in the in vitro-synthesized product. *Journal of Bacteriology*, 143, 293-301.

Chopra, I. (2013) The 2012 Garrod lecture: discovery of antibacterial drugs in the 21st century. *Journal of Antimicrobial Chemotherapy*, 68, 496-505.

Cohen, S.S. (1977) A strategy for the chemotherapy of infectious disease. *Science*, 197, 431-432.

Compton, C.L., Schmitz, K.R., Sauer, R.T. and Sello, J.K. (2013) Antibacterial activity of and resistance to small molecule inhibitors of the ClpP peptidase. *ACS Chemical Biology*, 8, 2669-2677.

Copeland, R.A (2010) Enzymes: a practical introduction to structure, mechanism, and data analysis. Second edition, *John Wiley and Sons, INC*.

Copeland, R.A (2005) Evaluation of enzyme inhibitors in drug discovery: A guide for medicinal chemists and pharmacologists. *John Wiley and Sons, INC*.

Crouvoisier, M., Auger, G., Blanot, D. and Mengin-Lecreulx, D. (2007) Role of the amino acid invariants in the active site of MurG as evaluated by site-directed mutagenesis. *Biochimie*, 89, 1498-1508.

Cruickshank, D.W.J. (1999) Remarks about protein structure precision. *Acta Crystallographica Section D*, 55, 583-601.

Culyba, M.J., Mo, C.Y. and Kohli, R.M. (2015) Targets for combating the evolution of acquired antibiotic resistance. *Biochemistry*, 54, 3579-3582.

Davis, D.C., Mohammad, H., Kyei-Baffour, K., Younis, W., Creemer, C.N., Seleem, M.N. and Dai, M. (2015) Discovery and characterization of aryl isonitriles as a new class of compounds versus methicillin- and vancomycin-resistant *Staphylococcus aureus*. *European Journal of Medicinal Chemistry*, 101, 384-390.

Dignam, J.D., Nada, S. and Chaires, J.B. (2003) Thermodynamic characterization of the binding of nucleotides to glycyl-tRNA synthetase. *Biochemistry*, 42, 5333-5340.

Dinescu, A., Cundari, T.R., Bhansali, V.S., Luo, J.L. and Anderson, M.E. (2004) Function of conserved residues of human glutathione synthetase: implications for the ATP-grasp enzymes. *Journal of Biological Chemistry*, 279, 22412-22421.

Doan, T.T., Kim, J.K., Ngo, H.P., Tran, H.T., Cha, S.S., Min Chung, K., Huynh, K.H., Ahn, Y.J. and Kang, L.W. (2014) Crystal structures of D-alanine-D-alanine ligase from *Xanthomonas oryzae* pv. *oryzae* alone and in complex with nucleotides. *Archives of Biochemistry and Biophysics*, 545, 92-99.

Dobson, R.C., Valegård, K. and Gerrard, J.A. (2004) The crystal structure of three site-directed mutants of *Escherichia coli* dihydrodipicolinate synthase: further evidence for a catalytic triad. *Journal of Molecular Biology*, 338, 329-339.

Done, H.Y., Venkatesan, A.K. and Halden, R.U. (2015) Does the recent growth of aquaculture create antibiotic resistance threats different from those associated with land animal production in agriculture? *The APPS Journal*, 17, 513-524.

Doyle, M.A., Gasser, R.B., Woodcroft, B.J., Hall, R.S. and Ralph, S.A. (2010) Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. *BMC Genomics*, 11, online.

Dulic, M., Cvetesic, N., Perona, J.J. and Gruic-Sovulj, I. (2010) Partitioning of tRNA-dependent editing between pre- and post-transfer pathways in class I aminoacyl-tRNA synthetases. *Journal of Biological Chemistry*, 285, 23799-23809.

Dyawanapelly, S., Ghodke, S.B., Vishwanathan, R., Dandekar, P. and Jain, R. (2014) RNA interference-based therapeutics: molecular platforms for infectious diseases. *Journal of Biomedical Nanotechnology*, 10, 1998-2037.

Ellsworth, B.A., Tom, N.J. and Bartlett, P.A. (1996) Synthesis and evaluation of inhibitors of bacterial D-alanine:D-alanine ligases. *Chemistry and Biology*, 3, 37-44.

Emptage, R.P., Daughtry, K.D., Pemble, C.W.-IV and Raetz, C.R. (2012) Crystal structure of LpxK, the 4'-kinase of lipid a biosynthesis and atypical P-loop kinase functioning at the membrane interface. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 12956-12961.

Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica Section D*, 66, 486-501.

Endrizzi, J.A., Kim, H., Anderson, P.M. and Baldwin, E.P. (2005) Mechanisms of product feedback regulation and drug resistance in cytidine triphosphate synthetases from the structure of a CTP-inhibited complex. *Biochemistry*, 44, 13491-13499.

Engh, R.A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica Section A*, 47, 392-400.

Erlanson, D.A. (2006) Fragment-based lead discovery: a chemical update. *Current Opinion in Biotechnology*, 17, 643-652.

Errington, J., Daniel, R.A. and Scheffers, D.J. (2003) Cytokinesis in bacteria. *Microbiology and Molecular Biology Reviews*, 67, 52-65.

Evans, P.R. (2005) Scaling and assessment of data quality. *Acta Crystallographica Section D*, 62, 72-82.

Evans, P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D*, 67, 282-292.

Fair, R.J. and Tor, Y. (2014) Antibiotics and bacterial resistance in the 21st century. *Perspectives in Medicinal Chemistry*, 6, 25-64.

Fan, C., Park, II-S., Walsh, C.T. and Knox, J.R. (1997) D-alanine:D-alanine ligase: phosphonate and phosphinate intermediates with wild type and the Y216F mutant. *Biochemistry*, 36, 2531-2538.

Farrow III, J.M. and Pesci, E.C. (2007) Two distinct pathways supply anthranilate as a precursor of the *Pseudomonas* quinolone signal. *Journal of Bacteriology*, 189, 3425-3433.

Fawaz, M.V., Topper, M.E. and Firestine, S.M. (2011) The ATP-grasp enzymes. *Bioorganic Chemistry*, 39, 189-191.

Fischbach, M.A. (2011) Combination therapies for combating antimicrobial resistance. *Current Opinion in Microbiology*, 14, 519-523.

Franklin, M.C., Cheung, J., Rudolph, M.J., Burshteyn, F., Cassidy, M., Gary, E., Hillerich, B., Yao, Z.K., Carlier, P.R., Totrov, M. and Love, J.D. (2015) Structural genomics for drug design against the pathogen *Coxiella burnetii*. *Proteins: Structure, Function, and Bioinformatics*, early view, online.

Franzini, R.M., Neri, D. and Scheuermann, J. (2014) DNA-encoded chemical libraries: advancing beyond conventional small-molecule libraries. *Accounts of Chemical Research*, 47, 1247-1255.

Frearson, J.A., Wyatt, P.G., Gilbert, I.H. and Fairlamb, A.H. (2007) Target assessment for antiparasitic drug discovery. Trends in Parasitology, 23, 589-595.

Frees, D., Gerth, U. and Ingmer, H. (2014) Clp chaperones and proteases are central in stress survival, virulence and antibiotic resistance of *Staphylococcus aureus*. *International Journal of Medical Microbiology*, 304, 142-149.

Freiberg, C. and Brötz-Oesterhelt, H. (2005) Functional genomics in antibacterial drug discovery. *Drug Discovery Today,* 10, 927-935.

Freiberg, C., Brunner, N.A., Schiffer, G., Lampe, T., Pohlmann, J., Brands, M., Raabe, M., Häbich, D. and Ziegelbauer, K. (20004) Identification and characterization of the first class of potent bacterial acetyl-CoA carboxylase inhibitors with antibacterial activity. *Journal of Biological Chemistry*, 279, 26066-26073.

Fyfe, P.K., Oza, S.L., Fairlamb, A.H. and Hunter, W.N. (2008) *Leishmania trypanothione* synthetase-amidase structure reveals a basis for regulation of conflicting synthetic and hydrolytic activities. *Journal of Biological Chemistry*, 283, 17672-17680.

Fyfe, P.K., Rao, V.A., Zemla, Z., Cameron, S. and Hunter, W.N. (2009) Specificity and mechanism of *Acinetobacter baumanii nicotinamidase*: implications for activation of the front-line tuberculosis drug pyrazinamide. *Angewandte Chemie International Edition*, 48, 9176-9179.

Gabanyi, M.J., Adams, P.D., Arnold, K., Bordoli, L., Carter, L.G., Flippen-Andersen, J., Gifford, L., Haas, J., Kouranov, A., McLaughlin, W.A., Micallef, D.I., Minor,

W., Shah, R., Schwede, T., Tao, Y.P., Westbrook, J.D., Zimmerman, M. and Berman, H.M. (2011) The Structural Biology Knowledgebase: a portal to protein structures, sequences, functions, and methods. *Journal of Structure and Functional Genomics*, 12, 45-54.

Galimand, M., Carniel, E. and Courvalin, P. (2006) Resistance of *Yersinia pestis* to antimicrobial agents. *Antimicrobial Agents and Chemotherapy*, 50, 3233-3236.

Galivan, J., Inglese, J., McGuire, J.J., Nimec, Z. and Coward, J.K. (1985) gamma-Fluoromethotrexate: synthesis and biological activity of a potent inhibitor of dihydrofolate reductase with greatly diminished ability to form poly-gamma-glutamates. *Proceedings of the National Academy of Sciences of the United States of America*, 82, 2598-2602.

Garrow, T.A., Admon, A. and Shane, B. (1992) Expression cloning of a human cDNA encoding folylpoly(gamma-glutamate) synthetase and determination of its primary structure. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 9151-9155.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31, 3784-3788.

Geiger, S.R., Böttcher, T., Sieber, S.A. and Cramer, P. (2011) A conformational switch underlies ClpP protease function. *Angewandte Chemie International Edition*, 50, 5749-5752.

Gelpi, A., Gilbertson, A. and Tucker, J.D. (2015) Magic bullet: Paul Ehrlich, Salvarsan and the birth of venereology. *Sexually Transmitted Infections*, 91, 68-69.

Ghanei, H., Abeyrathne, P.D. and Lam, J.S. (2007) Biochemical characterization of MsbA from *Pseudomonas aeruginosa*. *Journal of Biological Chemistry*, 282, 26939-26947.

Gillner, D., Armoush, N., Holz, R.C. and Becker, D.P. (2009) Inhibitors of bacterial N-succinyl-L,L-diaminopimelic acid desuccinylase (DapE) and demonstration of in vitro antimicrobial activity. *Bioorganic and Medicinal Chemistry Letters*, 19, 6350-6352.

Goto, M., Omi, R., Nakagawa, N., Miyahara, I. and Hirotsu, K. (2004) Crystal structures of CTP synthetase reveal ATP, UTP, and glutamine binding sites. *Structure*, 12, 1413-1423.

Green, O.M., McKenzie, A.R., Shapiro, A.B., Otterbein, L., Ni, H., Patten, A., Stokes, S., Albert, R., Kawatkar, S. and Breed, J. (2012) Inhibitors of acetyltransferase domain of N-acetylglucosamine-1-phosphate-uridyltransferase/glucosamine-1-phosphate-acetyltransferase (GlmU). Part 1: Hit to lead evaluation of a novel arylsulfonamide series. *Bioorganic and Medicinal Chemistry Letters*, 22, 1510-1519.

Greenfield, R.A., Drevets, D.A., Machado, L.J., Voskuhl, G.W., Cornea, P. and Bronze, M.S. (2002) Bacterial pathogens as biological weapons and agents of bioterrorism. *The American Journal of the Medical Sciences*, 323, 299-315.

Guo, R.T., Kuo, C.J., Chou, C.C., Ko, T.P., Shr, H.L., Liang, P.H. and Wang, A.H. (2004) Crystal structure of octaprenyl pyrophosphate synthase from hyperthermophilic *Thermotoga maritima* and mechanism of product chain length determination. *Journal of Biological Chemistry*, 279, 4903-4912.

Gupta, A., Mir, S.S., Saqib, U., Biswas, S., Vaishya, S., Srivastava, K., Siddiqi, M.I. and Habib, S. (2013) The effect of fusidic acid on *Plasmodium falciparum*

elongation factor G (EF-G). *Molecular and Biochemical Parasitology*, 192, 39-48.

Gwynn, M.N., Portnoy, A., Rittenhouse, S.F. and Payne, D.J. (2010) Challenges of antibacterial discovery revisited. *Annals of the New York Academy of Sciences*, 1213, 5-19.

Haertlé, T., Wohlrab, F. and Guschlbauer, W. (1979) Thymidylate synthetase from *Escherichia coli* K12. Purification, and dependence of kinetic properties on sugar conformation and size of the 2' substituent. *European Journal of Biochemistry*, 102, 223-230.

Hamamoto, H., Urai, M., Ishii, K., Yasukawa, J., Paudel, A., Murai, M., Kaji, T., Kuranaga, T., Hamase, K., Katsu, T., Su, J., Adachi, T., Uchida, R., Tomoda, H.5, Yamada, M., Souma, M., Kurihara, H., Inoue, M. and Sekimizu, K. (2015) Lysocin E is a new antibiotic that targets menaquinone in the bacterial membrane. *Nature Chemical Biology*, 11, 121-133.

Han, Q., Robinson, H. and Li, J. (2012) Biochemical identification and crystal structure of kynurenine formamidase from *Drosophila melanogaster*. *Biochemical Journal*, 446, 253-260.

Harding, M.M. (2001) Geometry of metal-ligand interactions in proteins. *Acta Crystallographica Section D*, 57, 401-411.

Harmark, K., Anborgh, P.H., Merola, M., Clark, B.F. and Parmeggiani, A. (1992) Substitution of aspartic acid-80, a residue involved in coordination of magnesium, weakens the GTP binding and strongly enhances the GTPase of the G domain of elongation factor Tu. *Biochemistry*, 31, 7367-7372.

Hasegawa, H. and Holm, L. (2009) Advances and pitfalls of protein structural alignment. *Current Opinion in Structural Biology*, 19, 341-348.

Hausner, T.P., Geigenmüller, U. and Nierhaus, K.H. (1988) The allosteric three-site model for the ribosomal elongation cycle. New insights into the inhibition mechanisms of aminoglycosides, thiostrepton, and viomycin. *Journal of Biological Chemistry*, 263, 13103-13111.

Haydon, D.J., Stokes, N.R., Ure, R., Galbraith, G., Bennett, J.M., Brown, D.R., Baker, P.J., Barynin, V.V., Rice, D.W., Sedelnikova, S.E., Heal, J.R., Sheridan, J.M., Aiwale, S.T., Chauhan, P.K., Srivastava, A., Taneja, A., Collins, I., Errington, J. and Czaplewski, L.G. (2008) An inhibitor of FtsZ with potent and selective anti-staphylococcal activity. *Science*, 321, 1673-1675.

He, X. and Reynolds, K.A. (2002) Purification, characterization, and identification of novel inhibitors of the beta-ketoacyl-acyl carrier protein synthase III (FabH) from *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 46, 1310-1308.

He, L., Zhang, L., Liu, X., Li, X., Zheng, M., Li, H., Yu, K., Chen, K., Shen, X., Jiang, H. and Liu, H. (2009) Discovering potent inhibitors against the beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ) of *Helicobacter pylori*: structure-based design, synthesis, bioassay, and crystal structure determination. *Journal of Medicinal Chemistry*, 52, 2465-2481

Heath, R.J. and Rock, C.O. (1996) Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. *Journal of Biological Chemistry*, 271, 27795-27801.

Hernick, M., Gattis, S.G., Penner-Hahn, J.E. and Fierke, C.A. (2010) Activation of *Escherichia coli* UDP-3-O-[(R)-3-hydroxymyristoyl]-N-acetylglucosamine

deacetylase by Fe^{2+} yields a more efficient enzyme with altered ligand affinity. *Biochemistry*, 49, 2246-2255.

Hiltunen, J.K., Schonauer, M.S., Autio, K.J., Mittelmeier, T.M., Kastaniotis, A.J. and Dieckmann, C.L. (2009) Mitochondrial fatty acid synthesis type II: more than just fatty acids. *Journal of Biological Chemistry*, 284, 9011-9015.

Hirashima, M., Hayakawa, T. and Koike, M. (1967) Mammalian alpha-keto acid dehydrogenase complexes. II. An improved procedure for the preparation of 2-oxoglutarate dehydrogenase complex from pig heart muscle. *Journal of Biological Chemistry*, 242, 902-907.

Hirsch, A.K., Lauw, S., Gersbach, P., Schweizer, W.B., Rohdich, F., Eisenreich, W., Bacher, A. and Diederich, F. (2007) Nonphosphate inhibitors of IspE protein, a kinase in the non-mevalonate pathway for isoprenoid biosynthesis and a potential target for antimalarial therapy. *ChemMedChem*, 2, 806-810.

Hofer, A., Steverding, D., Chabes, A., Brun, R. and Thelander, L. (2001) *Trypanosoma brucei* CTP synthetase: a target for the treatment of African sleeping sickness. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 6412-6416.

Hofmann, S.G., Meuret, A.E., Smits, J.A., Simon, N.M., Pollack, M.H., Eisenmenger, K., Shiekh, M. and Otto, M.W. (2006) Augmentation of exposure therapy with D-cycloserine for social anxiety disorder. *Archives of General Psychiatry*, 63, 298-304.

Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research*, 38, 545-549.

Holm, L. and Sander, C. (1997) An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins: Structure, Function and Bioinformatics*, 28, 72-82.

Houser, R.M. and Olson, R.E. (1977) 5-demethylubiquinone-9-methyltransferase from rat liver mitochondria. Characterization, localization, and solubilization. *Journal of Biological Chemistry*, 252, 4017-4021.

Huang, M. and Graves, L.M. (2002) *De novo* synthesis of pyrimidine nucleotides; emerging interfaces with signal transduction pathways. *Cellular and Molecular Life Sciences*, 60, 321-336.

Huggins, D.J., Sherman, W. and Tidor, B. (2012) Rational approaches to improving selectivity in drug design. *Journal of Medicinal Chemistry*, 55, 142-144.

Ingvarsson, H., Maté, M.J., Högbom, M., Portnoï, D., Benaroudj, N., Alzari, P.M., Ortiz-Lombardía, M. and Unge, T. (2007) Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of *Mycobacterium tuberculosis* ClpP1. *Acta Crystallographica Section D*, 63, 249-259.

Iscla, I., Wray, R., Blount, P., Larkins-Ford, J., Conery, A.L., Ausubel, F.M., Ramu, S., Kavanagh, A., Huang, J.X., Blaskovich, M.A., Cooper, M.A., Obregon-Henao, A., Orme, I., Tjandra, E.S., Stroeher, U.H., Brown, M.H., Macardle, C., van Holst, N., Ling Tong, C., Slattery, A.D., Gibson, C.T., Raston, C.L. and Boulos, R.A. (2015) A new antibiotic with potent activity targets MscL. *Journal of Antibiotics*, 68, 453-462.

Iwashima, A., Nosaka, K., Nishimura, H. and Kimura, Y. (1986) Some properties of a *Saccharomyces cerevisiae* mutant resistant to 2-amino-4-methyl-5-beta-hydroxyethylthiazole. *Journal of General Microbiology*, 132, 1541-1546.

Jaberi, E., Chitsazian, F., Ali, Shahidi, G., Rohani, M., Sina, F., Safari, I., Malakouti Nejad, M., Houshmand, M., Klotzle, B. and Elahi, E. (2013) The novel mutation p.Asp251Asn in the β-subunit of succinate-CoA ligase causes encephalomyopathy and elevated succinylcarnitine. *Journal of Human Genetics*, 58, 526-530.

Jacobs, M.A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., Will, O., Kaul, R., Raymond, C., Levy, R., Chun-Rong, L., Guenthner, D., Bovee, D., Olson, M.V. and Manoil, C. (2003) Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 14339-14344.

Jensen Jr, K.A., Ryan, Z.C., Vanden, Wymelenberg, A., Cullen, D. and Hammel, K.E. (2002) An NADH:quinone oxidoreductase active during biodegradation by the brown-rot basidiomycete *Gloeophyllum trabeum*. *Applied and Environmental Microbiology*, 68, 2699-2703.

Johnson, K.A. and Goody, R.S. (2011) The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper. *Biochemistry*, 50, 8264-82690.

Johanson, K., Hoang, T., Sheth, M. and Hyman, L.E. GRS1, a yeast tRNA synthetase with a role in mRNA 3' end formation. *Journal of Biological Chemistry*, 278, 35923-35930.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic acids research*, 36, 5-9.

Jolly, L., Ferrari, P., Blanot, D., van Heijenoort, J., Fassy, F. and Mengin-Lecreulx, D. (1999) Reaction mechanism of phosphoglucosamine mutase from *Escherichia coli*. *European Journal of Biochemistry*, 262, 202-210.

Joosten, R.P., Long, F., Murshudov, G.N. and Perrakis, A. (2014) The PDB_REDO server for macromolecular structure model optimization. *International Union of Crystallography Journal*, 1, 213-220.

Josefsen, L., Bohn, L., Sørensen, M.B. and Rasmussen, S.K. (2007) Characterization of a multifunctional inositol phosphate kinase from rice and barley belonging to the ATP-grasp superfamily. *Gene*, 397, 114-125.

Kabsch, W. (2010) XDS. *Acta Crystallographica Section D*, 66, 125-132.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2013) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, 42, 199-205.

Kantardjieff, K.A. and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Science*, 12, 1865-1871.

Katz, E., Brown, D. and Hitchcock, M. J. (1987) Arylformamidase from *Streptomyces parvulus*. *Methods in Enzymology*, 142, 225-234.

Kaul, G. and Pattan, G. (2011) MsbA ATP-binding cassette (ABC) transporter of E. coli: structure and possible flippase mechanism. *Indian Journal of Biochemistry and Biophysics*, 48, 7-13.

Kawate, H., Landis, D.M. and Loeb, L.A. (2002) Distribution of mutations in human thymidylate synthase yielding resistance to 5-fluorodeoxyuridine. *Journal of Biological Chemistry*, 277, 36304-36311.

Keedy, D.A., van den Bedem, H., Sivak, D.A., Petsko, G.A., Ringe, D., Wilson, M.A. and Fraser, J.S. (2014) Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR. *Structure*, 22, 899-910.

Keller, T.H., Pichota, A. and Yin, Z. (2006) A practical view of 'druggability'. *Current Opinion in Chemical Biology*, 10, 357-361.

Kiga, D., Sakamoto, K., Sato, S., Hirao, I. and Yokoyama, S. (2001) Shifted positioning of the anticodon nucleotide residues of amber suppressor tRNA species by *Escherichia coli* arginyl-tRNA synthetase. *European Journal of Biochemistry*, 268, 6207-6213.

Kilcawley, K.N., Wilkinson, M.G. and Fox, P.F. (2002) Determination of key enzyme activities in commercial peptidase and lipase preparations from microbial or animal sources. *Enzyme and Microbial Technology*, 31, 310-320.

Kim, D.Y. and Kim K.K. (2008) The structural basis for the activation and peptide recognition of bacterial ClpP. *Journal of Molecular Biology*, 379, 760-771.

Kingwell, K. (2015) Bacteriophage therapies re-enter clinical trials. *Nature Reviews Drug Discovery*, 14, 515-516.

Krasowski, A., Muthas, D., Sarkar, A., Schmitt, S. and Brenk, R. (2011) DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set. *Journal of Chemical Information and Modeling*, 51, 2829-2842.

Kresze, G.B. and Ronft H. (1981) Pyruvate dehydrogenase complex from baker's yeast. 1. Purification and some kinetic and regulatory properties. *European Journal of Biochemistry*, 119, 573-579.

Krissinel, E. and Henrick K. (2005). Detection of protein assemblies in crystals. *Computational Life Sciences: Lecture Notes in Computer Science* (*Springer)*, 3695, 163-174.

Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372, 774-797.

Koike, K., Suematsu, T. and Ehara, M. (2000) Cloning, overexpression and mutagenesis of cDNA encoding dihydrolipoamide succinyltransferase component of the porcine 2-oxoglutarate dehydrogenase complex. *European Journal of Biochemistry*, 267, 3005-3016.

Kovač, A., Majce, V., Lenarsic R., Bombek, S., Bostock, J.M., Chopra, I., Polanc, S. and Gobec, S. (2007) Diazenedicarboxamides as inhibitors of. *Bioorganic and Medicinal Chemistry Letters*, 17, 2047-2054.

Kurnasov, O., Goral, V., Colabroy, K., Gerdes, S., Anantha, S., Osterman, A. and Bedley, T.P. (2003) NAD biosynthesis: identification of the tryptophan to quinolinate pathway in bacteria. *Chemistry and Biology*, 10, 1195-1204.

Kurnasov, O., Jablonski, L., Polanuyer, B., Dorrestein, P., Begley, T. and Osterman, A. (2003) Aerobic tryptophan degradation pathway in bacteria: novel kynurenine formamidase. *FEMS Microbiology Letters*, 227, 219–227.

Kwon, Y.K., Lu, W., Melamud, E., Khanam, N., Bognar, A. and Rabinowitz, J.D. (2008) A domino effect in antifolate drug action in Escherichia coli. *Nature Chemical Biology*, 4, 602-608.

Kyuma, T., Kimura, S., Hanada, Y., Suzuki, T., Sekimizu, K. and Kaito, C. (2015) Ribosomal RNA methyltransferases contribute to *Staphylococcus aureus* virulence. *FEBS Journal*, 282, 2570-2584.

Lam, H., Oh, D.-C., Cava F., Takacs C.N., Clardy, J., de pedro, M.A. and Waldor, M.K. (2009) D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science*, 325, 1552-1555.

Landeka, I., Filipic-Rocak, S., Zinic, B. and Weygand-Durasevic, I. (2002) Characterization of yeast seryl-tRNA synthetase active site mutants with improved discrimination against substrate analogues. *Biochimica et Biophysica Acta*, 1480, 160-170.

Lapointe, J. and Söll, D. (1972) Glutamyl transfer ribonucleic acid synthetase of *Escherichia coli*. II. Interaction with intact glutamyl transfer ribonucleic acid. *Journal of Biological Chemistry*, 247, 4975-4981.

Lee, P.J., Bhonsle, J.B., Gaona, H.W., Huddler, D.P., Heady, T.N., Kreishman-Deitrick, M., Bhattacharjee, A., McCalmont, W.F., Gerena, L., Lopez-Sanchez, M., Roncal, N.E., Hudson, T.H., Johnson, J.D., Prigge, S.T. and Waters, N.C. (2006) Targeting the fatty acid biosynthesis enzyme, beta-ketoacyl-acyl carrier protein synthase III (PfKASIII), in the identification of novel antimalarial agents. *Journal of Medicinal Chemistry*, 52, 952-963.

Lee, B.-G., Kim, M.K. and Song, H.K. (2011) Structural insights into the conformational diversity of ClpP from *Bacillus subtilis*. *Molecules and Cells*, 32, 589-595.

Lee, M., Maher, M.L., Christopherson, R.I. and Guss, J.M. (2007) Kinetic and structural analysis of mutant *Escherichia coli* dihydroorotases: a flexible loop stabilizes the transition state. *Biochemistry*, 46, 10538-10550.

Lee, J.H., Na, Y., Song, H.E., Kim, D., Park, B.H., Rho, S.H., Im, Y.J., Kim, M.K., Kang, G.B., Lee, D.S. and Eom, S.H. (2006) Crystal structure of the apo form of D-alanine:D-alanine ligase (Ddl) from *Thermus caldophilus*: a basis for the substrate-induced conformational changes. *Proteins: Structure, Function, and Bioinformatics*, 64, 1078-1082.

Lee, B.-G., Park, E.Y., Lee, K.-E., Jeon, H., Sung, K.H., Paulsen, H., Rubsamen-Schaeff, H., Brotz-Oesterhelt, H. and Song, H.K. (2010) Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nature*, 17, 471-478.

Leslie, A.G.W and Powell, H.R. (2007) Evolving methods for macromolecular crystallography. *Springer*, 245, 41-51.

Li, D.H., Chung, Y.S., Gloyd, M., Joseph, E., Ghirlando, R., Wright, G.D., Cheng, Y.Q., Maurizi, M.R., Guarné, A. and Ortega, J. (2010) Acyldepsipeptide antibiotics induce the formation of a structured axial channel in ClpP: A model for the ClpX/ClpA-bound state of ClpP. *Chemistry and Biology*, 17, 959-969.

Li, X., Meng, X., Duan, H., Wang, L., Wang, S., Zhang, Y. and Qin, D. (2010) Original and efficient synthesis of D-cycloserine. *Archiv der Pharmazie: Chemistry in Life Sciences*, 8, 473-475.

Li, X.Z. and Nikaido, H. (2009) Efflux-mediated drug resistance in bacteria: an update. *Drugs*, 69, 1555-15623.

Liang, X., Lee, C.J., Zhao, J., Toone, E.J. and Zhou, P. (2013) Synthesis, structure, and antibiotic activity of aryl-substituted LpxC inhibitors. *Journal of Medicinal Chemistry*, 56, 6954-6966.

Liberati, N.T., Urbach, J.M., Miyata, S., Lee, D.G., Drenkard, E., Wu, G., Villanueva, J., Wei, T. and Ausubel, F.M. (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 2833-2838.

Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., Mueller, A., Schäberle, T.F., Hughes, D.E., Epstein, S., Jones, M., Lazarides, L., Steadman, V.A., Cohen, D.R., Felix, C.R., Fetterman, K.A., Millett, W.P., Nitti, A.G., Zullo, A.M., Chen, C. and Lewis, K. (2015) A new antibiotic kills pathogens without detectable resistance. *Nature*, 517, 455-459.

Liu, J.L. (2011) The enigmatic cytoophidium: compartmentation of CTP synthase via filament formation. *Bioessays*, 33, 159-164.

Liu, S., Chang, J.S., Herberg, J.T., Homg, M.-M., Tomich, P.K., Lin, A.H. and Marotti, K.R. (2006) Allosteric inhibition of *Staphylococcus aureus* D-alanine:D-alanine ligase revealed by crystallographic studies. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 15178-15183.

Liu, J., Ibba, M., Hong, K.W. and Söll, D. (1998) The terminal adenosine of tRNA(Gln) mediates tRNA-dependent amino acid recognition by glutaminyl-tRNA synthetase. *Biochemistry*, 37, 9836-9842.

Liu, H., Ritter, T.K., Sadamoto, R., Sears, P.S., Wu , M. and Wong, C.H. (2003) Acceptor specificity and inhibition of the bacterial cell-wall glycosyltransferase MurG. *ChemBioChem*, 4, 603-609.

Lo, M.C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M. and Ellestad, G. (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical Biochemistry*, 332, 153-159.

Lomovskaya, O. and Watkins, W. (2001) Inhibition of efflux pumps as a novel approach to combat drug resistance in bacteria. *Journal of Molecular Microbiology and Biotechnology*, 3, 225-236.

Løset, G.Å. and Sandlie, I. (2012) Next generation phage display by use of pVII and pIX as display scaffolds. *Methods*, 58, 40-46.

Lu, C. and Erickson, H.P. (1999) The straight and curved conformation of FtsZ protofilaments-evidence for rapid exchange of GTP into the curved protofilament. *Cell Structure and Function*, 24, 285-290.

Lu, Y., Xu, H. and Zhao, X. (2010) Crystal structure of the Apo form of D-Alanine:D-Alanine ligase (Ddl) from *Streptococcus mutans*. *Protein and Peptide Letters*, 17, 1053-1057.

Lunn, F.A. and Bearne, S.L. (2004) Alternative substrates for wild-type and L109A *E. coli* CTP synthases: kinetic evidence for a constricted ammonia tunnel. *European Journal of Biochemistry*, 271, 4204-4212.

Ma, S. and Ma, S. (2012) The development of FtsZ inhibitors as potential antibacterial agents. *ChemMedChem*, 7, 1161-1172.

Maddison, D.C. and Giorgini, F. (2015) The kynurenine pathway and neurodegenerative disease. *Seminars in Cell and Developmental Biology*, 40, 134-141.

Mathieu, M., Debousker, G., Vincent, S., Viviani, F., Bamas-Jacques, N. and Mikol, V. (2005) *Escherichia coli* FolC structure reveals an unexpected dihydrofolate binding site providing an attractive target for anti-microbial therapy. *Journal of Biological Chemistry*, 280, 18916-18922.

Maurizi, M.R., Clark, W.P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B. and Gottesman, S. (1990) Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of *Escherichia coli*. *Journal of Biological Chemistry*, 265, 12536-12545.

McAllister, L.A., Montgomery, J.I., Abramite, J.A., Reilly, U., Brown, M.F., Chen, J.M., Barham, R.A., Che, Y., Chung, S.W., Menard, C.A., Mitton-Fry, M., Mullins, L.M., Noe, M.C., O'Donnell, J.P., Oliver, R.M.-III, Penzien, J.B., Plummer, M., Price, L.M., Shanmugasundaram, V., Tomaras, A.P. and Uccello, D.P. (2012) Heterocyclic methylsulfone hydroxamic acid LpxC inhibitors as Gram-negative antibacterial agents. *Bioorganic and Medicinal Chemistry Letters*, 22, 6832-6838.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*, 40, 658-674.

Mengin-Lecreulx, D., Blanot, D. and van Heijenoort, J. (1994) Replacement of diaminopimelic acid by cystathionine or lanthionine in the peptidoglycan of *Escherichia coli*. *Journal of Bacteriology*, 176, 4321-4327.

Messmer, M., Blais, S.P., Balg, C., Chênevert, R., Grenier, L., Lagüe, P., Sauter, C., Sissler, M., Giegé, R., Lapointe, J. and Florentz, C. (2009) Peculiar inhibition of human mitochondrial aspartyl-tRNA synthetase by adenylate analogs. *Biochimie*, 91, 596-603.

Metlitskaya, A., Kazakov, T., Kommer, A., Pavlova, O., Praetorius-Ibba, M., Ibba, M., Krasheninnikov, I., Kolb, V., Khmel, I. and Severinov, K. (2006) Aspartyl-tRNA synthetase is the target of peptide nucleotide antibiotic Microcin C. *Journal of Biological Chemistry*, 281, 18033-18042.

Meziane-Cherif, D., Saul, F.A., Moubareck, C., Weber, P., Haouz, A., Courvalin, P. and Périchon, B. (2010) Molecular basis of vancomycin dependence in VanA-type *Staphylococcus aureus* VRSA-9. *Journal of Bacteriology*, 192, 5465-5471.

Monaghan, R.L. and Barrett, J.F. (2006) Antibacterial drug discovery--then, now and the genomics future. *Biochemical Pharmacology*, 71, 901-909.

Montgomery, J.I., Toogood, P.L., Hutchings, K.M., Liu, J., Narasimhan, L., Braden, T., Dermyer, M.R., Kulynych, A.D., Smith, Y.D., Warmus, J.S. and Taylor, C. (2009) Discovery and SAR of benzyl phenyl ethers as inhibitors of bacterial phenylalanyl-tRNA synthetase. *Bioorganic and Medicinal Chemistry Letters*, 19, 665-669

Mori, G., Chiarelli, L.R., Esposito, M., Makarov, V., Bellinzoni, M., Hartkoorn, R.C., Degiacomi, G., Boldrin, F., Ekins, S., de Jesus Lopes Ribeiro, A.L., Marino, L.B., Centárová, I., Svetlíková, Z., Blaško, J., Kazakova, E., Lepioshkin, A., Barilone, N., Zanoni, G., Porta, A., Fondi, M., Fani, R., Baulard, A.R., Mikušová, K., Alzari, P.M., Manganelli, R., de Carvalho, L.P., Riccardi, G., Cole, S.T. and Pasca, M.R. (2015) Thiophenecarboxamide derivatives activated by EthA kill *Mycobacterium tuberculosis* by inhibiting the CTP synthetase PyrG. *Chemistry and Biology*, 22, 917-927.

Mori, S., Kawai, S., Shi, F., Mikami, B. and Murata, K. (2005) Molecular conversion of NAD kinase to NADH kinase through single amino acid residue substitution. *Journal of Biological Chemistry*, 280, 24104-24112.

Moynié, L., Leckie, S.M., McMahon, S.A., Duthie, F.G., Koehnke, A., Taylor, J.W., Alphey, M.S., Brenk, R., Smith, A.D. and Naismith, J.H. (2013) Structural insights into the mechanism and inhibition of the β-hydroxydecanoyl-acyl carrier protein dehydratase from *Pseudomonas aeruginosa*. *Journal of Molecular Biology*, 425, 365-377.

Moynié, L., Schnell, R, McMahon, S.A., Sandalova, T., Boulkerou, W.A., Schmidberger, J.W., Alphey, M., Cukier, C., Duthie, F., Kopec, J., Liu, H.,

Jacewicz, A., Hunter, W.N., Naismith, J.H. and Schneider, G. (2013) The AEROPATH project targeting *Pseudomonas aeruginosa*: crystallographic studies for assessment of potential targets in early-stage drug discovery. *Acta Crystallographica Section F*, 69, 25-34.

Murata, T., Bognar, A.L., Hayashi, T., Ohnishi, M., Nakayama, K. and Terawaki, Y. (2000) Molecular analysis of the folC gene of *Pseudomonas aeruginosa*. *Microbiology and Immunology*, 44, 879-886.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, N.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D*, 67, 355-367.

Neuhaus, F.C. and Baddiley, J. (2003) A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in Gram-positive bacteria. *Microbiology and Molecular Biology reviews*, 67, 686-723.

Nishino, H. (1972) Biogenesis of cocarboxylase in *Escherichia coli*. Partial purification and some properties of thiamine monophosphate kinase. *Journal of Biochemistry*, 72, 1093-1100.

Nobrega, F.L., Costa, A.R., Kluskens, L.D. and Azeredo, J. (2015) Revisiting phage therapy: new applications for old resources. *Trends in Microbiology*, 23, 185-191.

Noree, C., Monfort, E., Shiau, A.K. and Wilhelm, J.E. (2014) Common regulatory control of CTP synthase enzyme activity and filament formation. *Molecular Biology of the Cell*, 25, 2282-2290.

Nzila, A. (2006) Inhibitors of *de novo* folate enzymes in *Plasmodium falciparum*. *Drug Discovery Today*, 11, 939-944.

Oliva, M.A., Trambaiolo, D. and Löwe, J. (2007) Structural insights into the conformational variability of FtsZ. *Journal of Molecular Biology*, 373, 1229-1242.

Ollivaux, C., Soyez, D. and Toullec, J.-Y. (2014). Biogenesis of D-amino acid containing peptides/proteins: where, when and how? *Journal of Peptide Science*, 20, 595-612.

O'Shea, R. and Moser, H.E. (2008) Physicochemical properties of antibacterial compounds: implications for drug discovery. *Journal of Medicinal Chemistry*, 51, 2871-2878.

Pace, H.C. and Brenner, C. (2001) The nitrilase superfamily: classification, structure and function. *Genome Biology*, 2, online.

Pearson, J.T., Dabrowski, M.J., Kung, I. and Atkins, W.M. (2005) The central loop of *Escherichia coli* glutamine synthetase is flexible and functionally passive. *Archives of Biochemistry and Biophysics*, 436, 397-405.

Parsons, W.H., Patchett, W.H., Bull, H.G., Schoen, W.R., Taub, D., Davidson, J., Combs, P.L., Springer, J.P., Gadebusch, H., Wiessberger, M.E.V, Mellin, T.N. and Busch, R.D. (1988) Phosphinic acid inhibitors of D-alanyl-D-alanine ligase. *Journal of Medicinal Chemistry*, 31, 1772-1778.

Payne, D.J., Gwynn, M.N., Holmes, D.J. and Pompliano, D.L. (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nature Reviews Drug Discovery*, 6, 29-40.

Pereira, D.S., Donald, L.J., Hosfield, D.J. and Duckworth, H.W. (1994) Active site mutants of *Escherichia coli* citrate synthase. Effects of mutations on catalytic and allosteric properties. *Journal of Biological Chemistry*, 269, 412-417.

Pegan, S.D., Tian, Y., Sershon, V. and Mesecar, A.D. (2010) A universal, fully automated high throughput screening assay for pyrophosphate and phosphate release from enzymatic reactions. *Combinatorial Chemistry and High Throughput Screening*, 13, 27-38.

Pohanka, M. and Skládal, P. (2009) *Bacillus anthracis*, *Francisella tularensis* and *Yersinia pestis*. The most important bacterial warfare agents. *Folia Microbiologica*, 54, 263-272.

Prosser, G.A. and de Carvalho, L.P.S. (2013). Kinetic mechanism and inhibition of *Mycobacterium tuberculosis* D-alanine:D-alanine ligase by the antibiotic D-cycloserine. *FEBS Journal*, 280, 1150-1166.

Prosser, G.A. and de Carvalho, L.P.S. (2013). Reinterpreting the Mechanism of Inhibition of *Mycobacterium tuberculosis* D-Alanine:D-Alanine Ligase by D-cycloserine. *Biochemistry*, 52, 7145-7149.

Putty, S., Rau, A., Jamindar, D., Pagano, P., Quinn, T.M., Schewiezer, H.P. and Gutheil, W.G. (2011) Characterization of D-boroAla as a novel broad-spectrum antibacterial agent targeting D-Ala-D-Ala Ligase. *Chemical Biology and Drug Design*, 78, 757-763.

Radika, K. and Raetz, C.R. (1988) Purification and properties of lipid A disaccharide synthase of *Escherichia coli*. *Journal of Biological Chemistry*, 263, 14859-14867.

Rahman, A., Ting, K., Cullen, K.M., Braidy, N., Brew, B.J. and Guillemin, G.J. (2009) The excitotoxin quinolinic acid induces tau phosphorylation in human neurons. *PLoS One*, 4, online.

Ramasastri, B.V. and Blakley, R.L. (1962) 5,10-Methylenetetrahydrofolic dehydrogenase from bakers' yeast: I. Partial purification and some properties. *Journal of Biological Chemistry*, 237, 1982-1998.

Ray, B.L., Painter, G. and Raetz, C.R. (1984) The biosynthesis of gram-negative endotoxin. Formation of lipid A disaccharides from monosaccharide precursors in extracts of *Escherichia coli*. *Journal of Biological Chemistry*, 259, 4852-4859.

Read, R.J. and Sussman, J.L. (2007) Evolving Methods for Macromolecular Crystallography, *Springer*.

Richard, S.B., Lillo, A.M., Tetzlaff, C.N., Bowman, M.E., Noel, J.P. and Cane, D.E. (2004) Kinetic analysis of *Escherichia coli* 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase, wild type and mutants, reveals roles of active site amino acids. *Biochemistry*, 43, 12189-97.

Riethmiller, S. (2005) From Atoxyl to Salvarsan: searching for the magic bullet. *Chemotherapy*, 51, 234-242.

Robbins, J.C. and Oxender, D.L. (1973) Transport systems for alanine, serine, and glycine in *Escherichia coli* K-12. *Journal of Bacteriology*, 116, 12-18.

Roland, S., Ferone, R., Harvey, R.J., Styles, V.L. and Morrison, R.W. (1979) The characteristics and significance of sulfonamides as substrates for *Escherichia coli* dihydropteroate synthase. *Journal of Biological Chemistry*, 254, 10337-10345.

Rosen, A.E., Brooks, B.S., Guth, E., Francklyn, C.S. and Musier-Forsyth, K. (2006) Evolutionary conservation of a functionally important backbone phosphate group critical for aminoacylation of histidine tRNAs. *RNA*, 12, 1315-1322.

Rossi, F., Han, Q., Li, J., Li, J. and Rizzi, M. (2004) Crystal structure of human kynurenine aminotransferase I. *Journal of Biological Chemistry*, 279, 50214-50220.

Roy, H., Ling, J., Irnov, M. and Ibba, M. (2004) Post-transfer editing in vitro and in vivo by the beta subunit of phenylalanyl-tRNA synthetase. *The EMBO journal*, 23, 4639-4648.

Ruan, B., Bovee, M.L., Sacher, M., Stathopoulos, C., Poralla, K., Francklyn, C.S. and Söll, D. (2005) A unique hydrophobic cluster near the active site contributes to differences in borrelidin inhibition among threonyl-tRNA synthetases. *Journal of Biological Chemistry*, 280, 571-577.

Ruan, B., Nakano, H., Tanaka, M., Mills, J.A., DeVito, J.A., Min, B., Low, K.B., Battista, J.R. and Söll, D. (2004) Cysteinyl-tRNA(Cys) formation in *Methanocaldococcus jannaschii*: the mechanism is still unknown. *Journal of Bacteriology*, 186, 8-14.

Saint Girons, I., Gilles, A.M., Margarita, D., Michelson, S., Monnot, M., Fermandjian, S., Danchin, A. and Bârzu, O. (1987) Structural and catalytic characteristics of *Escherichia coli* adenylate kinase. *Journal of Biological Chemistry*, 262, 622-629.

Sampei, G., Baba, S., Kanagawa, M., Yanai, H., Ishii, T., Kawai, H., Fukai, Y., Ebihara, A., Nakagawa, N. and Kawai, G. (2010) Crystal structures of glycinamide ribonucleotide synthetase, PurD, from thermophilic eubacteria. *Journal of Biochemistry*, 148, 429-438.

Sankaran, K. and Wu, H.C. (1995) Bacterial prolipoprotein signal peptidase. *Methods in Enzymology*, 248, 169-180.

Santoro, N., Brtva, T., Roest, S.V., Siegel, K. and Waldrop, G.L. (2006) A high-throughput screening assay for the carboxyltransferase subunit of acetyl-CoA carboxylase. *Analytical Biochemistry*, 354, 70-77.

Sasaki, S. (1989) Numerical tables of anomalous scattering factors calculated by the Cromer and Liberman's Method. *KEK Report 88-14*, National Laboratory for High Energy Physics, Tsukuba.

Satoh, Y., Kuratsu, M., Kobayashi, D. and Dairi, T. (2014) New gene responsible for para-aminobenzoate biosynthesis. *Journal of Bioscience and Bioengineering*, 117, 178-183.

Schenk, G., Mitic , N., Gahan, L.R., Ollis, D.L., McGeary, R.P. and Guddat, L.W. (2012) Binuclear metallohydrolases: complex mechanistic strategies for a simple chemical reaction. *Accounts of Chemical Research*, 45, 1593-1904.

Schmidt, A., Wu, H., MacKenzie, R.E., Chen, V.J., Bewly, J.R., Ray, J.E., Toth, J.E. and Cygler, M. (2000) Structures of three inhibitor complexes provide insight into the reaction mechanism of the human methylenetetrahydrofolate dehydrogenase/cyclohydrolase. *Biochemistry*, 39, 6325-6335.

Schneider-Yin, X., Ulbrichova, D., Mamet, R., Martasek, P., Marohnic, C.C., Goren, A., Minder, E.I. and Schoenfeld, N. (2008) Characterization of two missense variants in the hydroxymethylbilane synthase gene in the Israeli population, which differ in their associations with acute intermittent porphyria. *Molecular Genetics and Metabolism*, 94, 343-346.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. and Schomburg, D. (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Research*, 32, 431-433.

Schweizer, H.P. (2003) Efflux as a mechanism of resistance to antimicrobials in *Pseudomonas aeruginosa* and related bacteria: unanswered questions. *Genetics and Molecular Research*, 31, 48-62.

Seibert, C. M. and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry*, 44, 6383-6392.

Sgraja, T., Alphey, M.S., Ghilagaber, S., Marquez, R., Robertson, M.N., Hemmings, J.L., Lauw, S., Rohdich, F., Bacher, A., Eisenreich, W., Illarionova, V. and Hunter, W.N. (2008) Characterization of *Aquifex aeolicus* 4-diphosphocytidyl-2C-methyl-d-erythritol kinase - ligand recognition in a template for antimicrobial drug discovery. *FEBS Journal*, 275, 2779-2794.

Shapiro, A.B., Ross, P.L., Gao, N., Livchak, S., Kern, G., Yang, W., Andrews, B. and Thresher, J. (2013) A high-throughput-compatible fluorescence anisotropy-based assay for competitive inhibitors of *Escherichia coli* UDP-N-acetylglucosamine acyltransferase (LpxA). *Journal of Biomolecular Screening*, 18, 314-317.

Sheng, Y., Cross, J., Shen, Y., Smith, C.A. and Bognar, A.L. (2002) Mutation of an essential glutamate residue in folylpolyglutamate synthetase and activation of the enzyme by pteroate binding. *Archives of Biochemistry and Biophysics*, 402, 94-103.

Sheng, Y., Khanam, N., Tsaksis, Y., Shi, X.M., Lu, Q.S. and Bognar, A.L. (2008) Mutagenesis of folylpolyglutamate synthetase indicates that dihydropteroate and tetrahydrofolate bind to the same site. *Biochemistry*, 47, 2388-2396.

Sheridan, R.P., Maiorov, V.N., Holloway, M.K., Cornell, W.D. and Gao, Y.D. (2010) Drug-like density: a method of quantifying the "bindability" of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank. *Journal of Chemcial Information and Modeling*, 50, 2029-2040.

Shoichet, B.K. (2006) Interpreting steep dose-response curves in early inhibitor discovery. *Journal of Medicinal Chemistry*, 49, 7274-7277.

Short, R. and Posch, A. (2011) Stain-free approach for Western blotting. *Genetic Engineering and Biotechnology News*, 31, online.

Sievers, F., Wilm, A., Dineen, D.G., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, online.

Silver, L.L. (2011) Challenges of antibacterial discovery. *Clinical Microbiology Reviews*, 24, 71-109.

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I.A., Lesley, S.A. and Godzik, A. (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics,* 23, 3403-3405.

Smith, C.A., Cross, J.A., Bognar, A.L. and Sun, X. (2006) Mutation of Gly51 to serine in the P-loop of *Lactobacillus casei* folylpolyglutamate synthetase abolishes activity by altering the conformation of two adjacent loops. *Acta Crystallographica Section D*, 62, 548-558.

Smith, M.A., Kannangara, C.G., Grimm, B. and von Wettstein, D. (1991) Characterization of glutamate-1-semialdehyde aminotransferase of *Synechococcus*. Steady-state kinetic analysis. *European Journal of Biochemistry*, 202, 749-757.

Soriano, A., Radice, A.D., Herbitter, A.H., Langsdorf, E.F., Stafford, J.M., Chan, S., Wang, S., Liu, Y.H., Black, T.A. (2006) *Escherichia coli* acetyl-coenzyme A

carboxylase: characterization and development of a high-throughput assay. *Analytical Biochemistry*, 349, 268-276.

Sova, M., Cadez, G., Turk, S., Majce, V., Polanc, S., Batson, S., Lloyd, A.J., Roper, D.I., Fishwick, C.W. and Gobec, S. (2009) Design and synthesis of new hydroxyethylamines as inhibitors of D-alanyl-D-lactate ligase (VanA) and D-alanyl-D-alanine ligase (DdlB). *Bioorganic & Medicinal Chemistry Letters*, 19, 1376-1379.

Sowole, M.A., Alexopoulos, J.A., Cheng, Y.-Q., Ortega, J. and Konermann, L. J. (2013). Activation of ClpP protease by ADEP antibiotics: insights from hydrogen exchange mass spectrometry. *Journal Molecular Biology*, 425, 4508-4519.

Stathopoulos, C., Jacquin-Becker, C., Becker, H.D., Li, T., Ambrogelly, A., Longman, R. and Söll, D. (2001) *Methanococcus jannaschii* prolyl-cysteinyl-tRNA synthetase possesses overlapping amino acid binding sites. *Biochemistry*, 40, 46-52.

Stone, T.W. and Darlington, L.G. (2002) Endogenous kynurenines as targets for drug discovery and development. *Nature Reviews Drug Discovery*, 1, 609-620.

Stone, T.W., Stoy, N. and Darlington, L.G. (2013) An expanding range of targets for kynurenine metabolites of tryptophan. *Trends in Pharmacological Sciences*, 34, 136-143.

Strater, N., Sun, L., Kantrowitz, E.R. and Lipscomb, W. N. (1999) A bicarbonate ion as a general base in the mechanism of peptide hydrolysis by dizinc leucine aminopeptidase. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 11151-11156.

Strauss, E., Kinsland, C., Ge, Y., McLafferty, F.W. and Begley, T.P. (2001) Phosphopantothenoylcysteine synthetase from *Escherichia coli*. Identification and characterization of the last unidentified coenzyme A biosynthetic enzyme in bacteria. *Journal of Biological Chemistry*, 276, 13513-13516.

Su, C.C., Long, F., McDermott, G., Shafer, W.M. and Yu, E.W. (2008) Crystallization and preliminary X-ray diffraction analysis of the multidrug efflux transporter NorM from *Neisseria gonorrhoeae*. *Acta Crystallographica Section F*, 64, 289-292.

Sun, X., Bognar, A.L., Baker, E.N. and Smith, C.A. (1998) Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthetase. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 6647-6652.

Sun, Y.H., Cheng, Q., Tian, W.X. and Wu, X.D. (2008) A substitutive substrate for measurements of beta-ketoacyl reductases in two fatty acid synthase systems. *Journal of Biochemical and Biophysical Methods*, 70, 850-856.

Sun, X., Cross, J.A., Bognar, A.L., Baker, E.N. and Smith, C.A. (2001) Folate-binding triggers the activation of folylpolyglutamate synthetase. *Journal of Biological Chemistry*, 310, 1067-1078.

Sun, T., Hayakawa, K. and Fraser, M.E. (2011) ADP-Mg^{2+} bound to the ATP-grasp domain of ATP-citrate lyase. *Acta Crystallographica Section F*, 67, 1168-1172.

Sun, Y., Song, H., Li, J., Jiang, M., Li, Y., Zhou, J. and Guo, Z. (2012) Active site binding and catalytic role of bicarbonate in 1,4-dihydroxy-2-naphthoyl coenzyme A synthases from vitamin K biosynthetic pathways. *Biochemistry*, 51, 4580-4589.

Sunchu, B., Berg, L., Ward, H.E. and Lopper, M.E. (2012) Identification of a small molecule PriA helicase inhibitor. *Biochemistry*, 51, 10137-10146.

Sybesma, W., van den Born, E., Starrenburg, M., Mierau, I., Kleerebezem, M., de Vos, W.M. and Hugenholtz, J. (2003) Controlled modulation of folate polyglutamyl tail length by metabolic engineering of *Lactococcus lactis*. *Applied and Environmental Microbiology*, 69, 7101-7107.

Szyk, A. and Maurizi, M.R. (2006) Crystal structure at 1.9 Å of *E. coli* ClpP with a peptide covalently bound at the active site. *Journal of Structural Biology*, 156, 165-174.

Takeda, H., Toyooka, T., Ikeuchi, Y., Yokobori, S., Okadome, K., Takano, F., Oshima, T., Suzuki, T., Endo, Y. and Hori, H. (2006) The substrate specificity of tRNA (m1G37) methyltransferase (TrmD) from *Aquifex aeolicus*. *Genes to Cells*, 11, 1353-1365.

Tamma, P.D., Cosgrove, S.E. and Maragakis, L.L. (2012) Combination therapy for treatment of infections with gram-negative bacteria. *Clinical Microbiology Reviews*, 25, 450-470.

Tasdemir, D., Lack, G., Brun, R., Rüedi, P., Scapozza, L. and Perozzo, R. (2006) Inhibition of *Plasmodium falciparum* fatty acid biosynthesis: evaluation of FabG, FabZ, and FabI as drug targets for flavonoids. *Journal of Medicinal Chemistry*, 49, 3345-3353.

Taylor, J.C. and Markham, G.D. (2003) Conformational dynamics of the active site loop of S-adenosylmethionine synthetase illuminated by site-directed spin labeling. *Archives of Biochemistry and Biophysics*, 415, 164-171.

Terrak, M., Sauvage, E., Derouaux, A., Dehareng, D., Bouhss, A., Breukink, E., Jeanjean, S. and Nguyen-Distèche, M. (2008) Importance of the conserved residues in the peptidoglycan glycosyltransferase module of the class A penicillin-binding protein 1b of *Escherichia coli*. *Journal of Biological Chemistry*, 283, 28464-28470.

Thompson, D., Lazennec, C., Plateau, P. and Simonson, T. (2007) Ammonium scanning in an enzyme active site. The chiral specificity of aspartyl-tRNA synthetase. *Journal of Biological Chemistry*, 282, 30856-30868.

Tommasi, R., Brown, D.G., Walkup, G.K., Manchester, J.I. and Miller, A.A. (2015) ESKAPEing the labyrinth of antibacterial discovery. *Nature Reviews Drug Discovery*, 14, 529-542.

Tonhosolo, R., D'Alexandri, F.L., Genta, F.A., Wunderlich, G., Gozzo, F.C., Eberlin, M.N., Peres, V.J., Kimura, E.A. and Katzin, A.M. (2005) Identification, molecular cloning and functional characterization of an octaprenyl pyrophosphate synthase in intra-erythrocytic stages of *Plasmodium falciparum*. *Biochemical Journal*, 392, 117-126.

Torok, M.E., Chantratita, N. and Peacock, S.J. (2012) Bacterial gene loss as a mechanism for gain of antimicrobial resistance. *Current Opinion in Microbioogy*, 15, 583-587.

Touw, W.G., Baakman, C., Black, J., te Beek, T.A.H., Krieger, E., Joosten, R.P. and Vriend, G. (2015) A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, 43, 364-368.

Turk, S., Hrast, M., Sosic, I., Barreteau, H., Mengin-Lecreulx, D., Blanot, D. and Gobec, S. (2013) Biochemical characterization of MurF from *Streptococcus*

pneumoniae and the identification of a new MurF inhibitor through ligand-based virtual screening. *Acta Chimica Slovenica*, 60, 294-299.

Uda, N.R. and Creus, M. (2011) Selectivity of inhibition of *N*-succinyl-L,L-diaminopimelic acid desuccinylase in bacteria: the product of dapE-gene is not the target of L-captopril antimicrobial activity. *Bioinorganic Chemistry and Applications*, 2011, online.

Vagin, A. and Teplyakov, A. (2009) Molecular replacement with MOLREP. *Acta Crystallographica Section D*, 66, 22-25.

Valderas, M.W., Andi, B., Barrow, W.W. and Cook, P.F. (2008) Examination of intrinsic sulfonamide resistance in *Bacillus anthracis*: a novel assay for dihydropteroate synthase. *Biochimica et Biophysica Acta*, 1780, 848-853.

van Opijnen, T. and Camilli, A. (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology*, 11, 435-442.

Vavricka, C.J., Liu, Y., Kiyota, H., Sriwilaijaroen, N., Qi, J., Tanaka, K., Wu, Y., Li, Q., Li, Y., Yan, J., Suzuki, Y. and Gao, G.F. (2013) Influenza neuraminidase operates via a nucleophilic mechanism and can be targeted by covalent inhibitors. *Nature Communications*, 4, online.

Vineyard, D., Patterson-Ward, J. and Lee, I. (2006) Single-turnover kinetic experiments confirm the existence of high- and low-affinity ATPase sites in *Escherichia coli* Lon protease. *Biochemistry*, 45, 4602-4610.

Voet, D. and Voet, J.G. (2004) Biochemistry. *John Wiley and Sons, INC*, third edition.

Wang, P., Wang, Q., Yang, Y., Coward, J.K., Nzila, A., Sims, P.F. and Hyde, J.E. (2010) Characterisation of the bifunctional dihydrofolate synthase-

folylpolyglutamate synthase from *Plasmodium falciparum*; a potential novel target for antimalarial antifolate inhibition. *Molecular and Biochemical Parasitology*, 172, 41-51.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189-1191.

Wei, Y., Zhang, H., Gao, Z.Q., Wang, W.J., Shtykova, E.V., Xu, J.H., Liu, Q.S. and Dong, Y.H. (2012) Crystal and solution structures of methyltransferase RsmH provide basis for methylation of C1402 in 16S rRNA. *Journal of Structural Biology*, 179, 29-40.

Weichenberger, C.X., Afonine, P.V., Kantardjieff, K. and Rupp, B. (2015) The solvent component of macromolecular crystals. *Acta Crystallographica Section D*, 71, 1023-1038.

White, S.W., Zheng, J., Zhang, Y.-M. and Rock, C.O. (2005). The structural biology of type II fatty acid biosynthesis. *Annual Review of Biochemistry*, 74, 791-831.

White, T.A. and Kell, D.B. (2004) Comparative genomic assessment of novel broad-spectrum targets for antibacterial drugs. *Comparative and Functional Genomics*, 5, 304-327.

White, N.J. (2003) Melioidosis. *Lancet*, 361, 1715-1722.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A. and Wilson, K.S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D*, 67, 235-242.

Wogulis, M., Chew, E.R., Donohoue, P.D. and Wilson, D.K. (2008) Identification of formyl kynurenine formamidase and kynurenine aminotransferase from *Saccharomyces cerevisiae*. Using crystallographic, bioinformatic and biochemical evidence. *Biochemistry*, 47, 1608-1621.

Wu, C., Dunaway-Mariano, D. and Mariano, P.S. (2013) Design, synthesis, and evaluation of inhibitors of pyruvate phosphate dikinase. *Journal of Organic Chemistry*, 78, 1910-1922.

Wu, D., Zhang, L., Kong, Y., Du, J., Chen, S., Chen, J., Ding, J., Jiang, H. and Shen, H. (2008) Enzymatic characterization and crystal structure analysis of the D-alanine-D-alanine ligase from *Helicobacter pylori*. *Proteins: Structure, Function and Bioinformatics*, 72, 1148-1160.

Wyatt, P.G., Gilbert, I.H., Read, K.D. and Fairlamb, A.H. (2011) Target validation: linking target and chemical properties to desired product profile. *Current Topics in Medicinal Chemistry*, 11, 1275-1283.

Xu, M.G., Li, J., Du, X. and Wang, E.D. (2004) Groups on the side chain of T252 in *Escherichia coli* leucyl-tRNA synthetase are important for discrimination of amino acids and cell viability. *Biochemical and Biophysical Research Communications*, 318, 11-16.

Yao, J. and Rock, C.O. (2015) How bacterial pathogens eat host lipids: implications for the development of fatty acid synthesis therapeutics. *Journal of Biological Chemistry*, 290, 5940-5946.

Yoshida, T., Nasu, H., Namba, E., Ubukata, O. and Yamashita, M. (2012) Discovery of a compound that acts as a bacterial PyrG (CTP synthase) inhibitor. *Journal of Medical Microbiology*, 61, 1280-1285.

Young, P.G., Smith, C.A., Metcalf, P. and Baker, E.N. (2008) Structures of *Mycobacterium tuberculosis* folylpolyglutamate synthase complexed with ADP and AMPPCP. *Acta Crystallographica Section D*, 64, 745-753.

Yu, A.Y. and Houry, W.A. (2007) ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Letters*, 581, 3749-3757.

Yue, D., Maizels, N. and Weiner, A.M. (1996) CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*. *RNA*, 2, 895-908.

Zajdowicz, S.L.W., Jones-Carson, J., Velazquez-Torres, A., Jobling, M.G., Gill, R.E. and Holmes, R.K. (2011) Alanine racemase mutants of *Burkholderia pseudomallei* and *Burkholderia mallei* and use of Alanine racemase as a non-antibiotic-based selectable marker. *PLoS ONE*, 6, online.

Zawadzke, L.E., Bugg, T.D. and Walsh, C.T. (1991) Existence of two D-alanine:D-alanine ligases in *Escherichia coli*: cloning and sequencing of the ddlA gene and purification and characterization of the DdlA and DdlB enzymes. *Biochemistry*, 30, 1673-1682.

Zepeck, F., Gräwert, T., Kaiser, J., Schramek, N., Eisenreich, W., Bacher, A. and Rohdich, F. (2005) Biosynthesis of isoprenoids. purification and properties of IspG protein from *Escherichia coli*. *Journal of Organic Chemistry*, 70, 9168-9174.

Zhang, G. (2012) Protease Assays. Assay Guidance Manual. *Eli Lilly and Company and the National Center for Advancing Translational Sciences*, Bethesda.

Zhang, R., Ou, H.Y. and Zhang, C.T. (2004) DEG: a database of essential genes. *Nucleic Acids Research*, 32, 271-272.

Zhang, C.M., Perona, J.J., Ryu, K., Francklyn, C. and Hou, Y.M. (2006) Distinct kinetic mechanisms of the two classes of aminoacyl-tRNA synthetases. *Journal of Molecular Biology*, 361, 300-311.

Zhang, J., Ye, F., Lan, L., Jiang, H., Luo, C. and Yang, C.G. (2011) Structural switching of *Staphylococcus aureus* Clp protease: a key to understanding protease dynamics. *Journal of Biological Chemistry*, 286, 37590-37601.

Zhang, H.J., Zhu, D.D., Li, Z.L., Sun, J. and Zhu, H.L. (2011) Synthesis, molecular modeling and biological evaluation of β-ketoacyl-acyl carrier protein synthase III (FabH) as novel antibacterial agents. *Bioorganic and Medicinal Chemistry Letters*, 19, 4513-4519.

Zummo, F.P., Marineo, S., Pace, A., Civiletti, F. (2012) Tryptophan catabolism via kynurenine production in *Streptomyces coelicolor*: identification of three genes coding for the enzymes of tryptophan to anthranilate pathway. *Applied Microbiology and Biotechnology*, 94, 719-728.

PUBLICATION

Díaz-Sáez, L., Srikannathasan, V., Zoltner, M. and Hunter, W.N. (2014) Structures of bacterial kynurenine formamidase reveal a crowded binuclear zinc catalytic site primed to generate a potent nucleophile. *The Biochemical Journal*, 462, 581-589.

Biochem. J. (2014) **462**, 581–589 (Printed in Great Britain) doi:10.1042/BJ20140511

Structures of bacterial kynurenine formamidase reveal a crowded binuclear zinc catalytic site primed to generate a potent nucleophile

Laura DÍAZ-SÁEZ*, Velupillai SRIKANNATHASAN*, Martin ZOLTNER* and William N. HUNTER*[1]

*Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, U.K.

Tryptophan is an important precursor for chemical entities that ultimately support the biosynthesis of key metabolites. The second stage of tryptophan catabolism is catalysed by kynurenine formamidase, an enzyme that is different between eukaryotes and prokaryotes. In the present study, we characterize the catalytic properties and present the crystal structures of three bacterial kynurenine formamidases. The structures reveal a new amidase protein fold, a highly organized and distinctive binuclear Zn^{2+} catalytic centre in a confined, hydrophobic and relatively rigid active site. The structure of a complex with 2-aminoacetophenone delineates aspects of molecular recognition extending to the observation that the substrate itself may be conformationally restricted to assist binding in the confined space of the active site and for subsequent processing. The cations occupy a crowded environment, and, unlike most Zn^{2+}-dependent enzymes, there is little scope to increase co-ordination number during catalysis. We propose that the presence of a bridging water/hydroxide ligand in conjunction with the placement of an active site histidine supports a distinctive amidation mechanism.

Key words: amidase, binuclear metal site, kynurenine formamidase, tryptophan catabolism, X-ray crystallography, zinc enzyme.

INTRODUCTION

Tryptophan catabolism via the kynurenine pathway supports the biosynthesis of NAD^+, important neuroactive intermediates in eukaryotes, then anthranilate, quinolate and antibiotics in prokaryotes [1–4]. The proteins involved in tryptophan catabolism are in general highly conserved across species, but with the notable exception of the second enzyme in the pathway: kynurenine formamidase. This enzyme catalyses the conversion of *N*-formyl-L-kynurenine (NFK) into formate and L-kynurenine [2].

In eukaryotes, the enzyme is termed kynurenine formamidase (KFase) and displays relatively low sequence conservation, although the structure itself is well maintained. This is evident from studies with *Sacharyomyces cerevisiae* and *Drosophila melanogaster* KFase, which, although they only share 10% sequence identity, are structurally conserved α/α hydrolases [5,6]. The KFase fold is dominated by a core eight-stranded α-sheet decorated on either side with a number of α-helices. This creates a well-defined cavity at one end of the sheet, leading to a catalytic triad consisting of a serine, histidine and aspartate combination. Overall, there are close structural similarities between KFase and carboxylesterases, but also intriguingly to another enzyme involved in kynurenine biology, namely a 5′-pyridoxyl phosphate-dependent kynurenine aminotransferase [5,6].

The chemistry behind an amidation reaction involves either acidic or basic hydrolysis. It is the latter that applies to these metal-free amidases, which exploit the properties of reactive cysteine or serine residues to create covalent acyl-intermediates during catalysis. Cysteine-dependent amidases are exemplified by trypanothione synthetase amidase [7], nicotinamidase [8] and the nitrilases [9], whereas the eukaryotic KFases from *S. cerevisiae* and *D. melanogaster* present examples of the serine-dependent enzymes with a mechanism that correlates with known amidases [5,6]. In KFase, an ordered mechanism results when the catalytic triad positions histidine to activate serine into a nucleophile that attacks the substrate carbonyl carbon. Subsequent engagement of a water molecule then allows for completion of the reaction.

Our attention was drawn to the prokaryotic KFase KynB (EC 3.5.1.9). Analysis of KynB sequences indicated that they are unrelated to KFase and indeed do not match any known amidase or amidohydrolase. Clearly, KynB is different and so, intrigued, we sought to investigate the structure–activity relationships in this distinctive enzyme. We now report the preparation of efficient recombinant expression systems for KynB from three bacterial sources and their purification. The catalytic properties of the enzymes are characterized and tryptophan fluorescence is used to investigate ligand binding. High-resolution crystal structures have been determined that allow us to describe the architecture of the enzyme extending to the details of the active site, to address metal ion identification and to investigate key features of molecular recognition at the catalytic centre. Taken together, our structural data then allow us to propose a plausible and distinctive mechanism for KynB.

EXPERIMENTAL PROCEDURES

Expression plasmid preparation

Genes encoding KynB in *Pseudomonas aeruginosa* (*Pa*KynB, UniProt: Q9I234), *Burkholderia cenocepacia* (*Bc*KynB, UniProt: B4E9I9) and *Bacillus anthracis* (*Ba*KynB, UniProt: Q81PP9) were purchased (Genscript) having been codon optimized for expression in *Escherichia coli* K12. *Pa*KynB and *Bc*KynB encoding genes were cloned into the NdeI/XhoI and *Ba*KynB into the NdeI/BamHI site of a modified pET15b vector (Novagen)

Abbreviations: ACMSD, α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase; *Ba*KynB, *Bacillus anthracis* KynB; *Bc*KynB *Burkholderia cenocepacia* KynB; CV, column volume; *Pa*KynB, *Pseudomonas aeruginosa* KynB; KFase, kynurenine formamidase; NCS, non-crystallographic symmetry; NFK, *N*-formyl-L-kynurenine; SEC, size-exclusion chromatography; TEV, tobacco etch virus; XANES, X-ray absorption near-edge structure.

[1] To whom correspondence should be addressed (email w.n.hunter@dundee.ac.uk).

Atomic co-ordinates and structure factors have been deposited in the PDB under codes 4COG, 4COB, 4CO9 and 4CZ1.

to create plasmids that produce an N-terminal hexahistidine-tagged (His-tagged) protein and a tobacco etch virus (TEV) protease cleavage site. All constructs were sequenced to check their integrity.

Expression and protein purification

Gene expression and purification of the proteins started with freshly transformed *E. coli* BL21(DE3)pLysS in the case of *Pa*KynB and *Bc*KynB systems, and *E. coli* BL21(DE3) for *Ba*KynB. Bacteria were cultured at 37 °C in 10 ml of LB broth containing 50 μg· ml^{-1} carbenicillin and, for *Pa*KynB and *Bc*KynB cultures, 20 μg· ml^{-1} chloramphenicol was also present. These cultures were used as inoculum for 1 litre of LB/antibiotic mixtures. The bacteria were cultured at 37 °C until an attenuance of 0.6–0.8 at $\lambda = 600$ nm was achieved. The temperature was lowered to 16 °C in the case of *Pa*KynB and *Bc*KynB, and to 20 °C for *Ba*KynB. Gene expression was induced with 1 mM IPTG and cultures were incubated overnight. Cells were then harvested by centrifugation (4000 ***g*** at 4 °C for 10 min). The *Pa*KynB and *Bc*KynB cultures were resuspended in buffer A1 (25 mM Tris/HCl, pH 7.5, and 100 mM NaCl), and *Ba*KynB in buffer A2 (20 mM Tris/HCl, pH 7.4, and 200 mM NaCl) with the addition of complete EDTA-free protease inhibitor cocktail tablets from Roche. The cells were disrupted using a sonicator in the case of *Pa*KynB and *Bc*KynB, and a French press for *Ba*KynB, and homogenates were centrifuged at 40000 ***g*** for 30 min at 4 °C. The resulting supernatants were passed through a 0.45 μm filter, and samples containing *Pa*KynB and *Bc*KynB were adjusted to contain 5 mM imidazole. Proteins were loaded on a 5-ml HisTrap HP column (GE Healthcare). After a 10-column volume (CV) washing step using 90 % buffers A1 or A2 and 10 % of buffer B1 (25 mM Tris/HCl, pH 7.5, 100 mM NaCl and 0.5 M imidazole) for *Pa*KynB and *Bc*KynB, or buffer B2 (20 mM Tris–HCl pH 7.4, 200 mM NaCl, 0.8 M imidazole) for *Ba*KynB, the recombinant proteins were eluted applying a linear imidazole gradient (50–250 mM for *Pa*KynB and *Bc*KynB, and 80–400 mM for *Ba*KynB over 20 CV). The eluted proteins were dialysed into buffer A1 or A2 and incubated overnight with 1 mg of His-tagged TEV protease per 20 mg of protein at 4 °C, then applied to a HisTrap HP column equilibrated with buffer A1 or A2 to remove the TEV protease, cleaved peptide and non-cleaved material. Initial experiments indicated that the His-tag of *Pa*KynB was not cleaved efficiently by TEV protease, and protein aggregation/precipitation was pronounced. We left the His-tag in place for further experiments with that protein. The enzyme assay (see later) indicated that the His-tagged *Pa*KynB possessed a level of activity comparable with that of other samples. Fractions containing the proteins were collected, concentrated and applied to a size-exclusion chromatography (SEC) column (HR 16/60, Superdex75 prep grade, GE Healthcare, CV = 120 ml) equilibrated with buffer C (10 mM NaH_2PO_4/Na_2HPO_4, pH 7.8, 20 mM NaCl and 0.5 mM Tris/HCl) in the case of *Pa*KynB and *Bc*KynB, and buffer A1 for *Ba*KynB. The SEC columns had been calibrated with molecular mass standards (thyroglobulin, 670 kDa; γ-globulin, 158 kDa; serum albumin, 67 kDa; ovalbumin, 44 kDa; myoglobin, 17 kDa; vitamin B_{12}, 1 kDa). Protein purity and molecular weight were assessed by SDS/PAGE and MALDI–TOF-MS analyses performed at the University of Dundee 'Fingerprints' Proteomics Facility using an Applied Biosystems Voyager DE-STR spectrometer. Protein quaternary structure was investigated by native PAGE and SEC.

Protein concentration determination

Protein concentrations were measured by absorbance at 280 nm using the predicted molar absorption coefficients by ExPASy bioinformatics resource portal ProtParam [10]:

$$\varepsilon(Pa\text{KynB}) = 28\,210\ \text{M}^{-1} \cdot \text{cm}^{-1}$$

$$\varepsilon(Bc\text{KynB}) = 28\,210\ \text{M}^{-1} \cdot \text{cm}^{-1}$$

$$\varepsilon(Ba\text{KynB}) = 20\,970\ \text{M}^{-1} \cdot \text{cm}^{-1}$$

Enzymatic assay and determination of catalytic parameters

A spectrophotometric assay was developed to obtain K_m, V_{max} and k_{cat} values for KynB. The assay is based on the increase in absorbance due to L-kynurenine formation [11,12] which is detected by measuring the absorbance at 365 nm with a UV-2450 Shimadzu spectrophotometer over a period of 160 s. The reactions were performed in triplicate using 1 ml final volume, at 25 °C, using 0.025, 0.15, 0.25, 0.5, 0.7, 1, 1.5, 2 and 3 mM NFK substrate (Dalton Pharma Services), and 500 ng of enzyme (20 nM). The buffer used contained 0.1 M NaH_2PO_4/Na_2HPO_4, pH 7.4, and 20 μM $ZnCl_2$. Calculations of the initial velocity for each substrate concentration were made by fitting a linear equation to the data of the increase in L-kynurenine concentration during 1 min, and data were analysed using the Michaelis–Menten equation (SigmaPlot, Systat Software). L-Kynurenine concentration was calculated using the Beer–Lambert law and ε_{365} of 4220 $\text{M}^{-1} \cdot \text{cm}^{-1}$ [12].

Crystallization, diffraction and structure determinations

*Pa*KynB and *Bc*KynB were concentrated to 7.5 mg · ml^{-1} in buffer A1 and *Ba*KynB to 4 mg· ml^{-1} in buffer A2 to provide a stock solution for crystallization experiments. The first crystallization trials used commercial screening sets from Molecular Dimensions and Qiagen and were set up in 96-well sitting drop plates with a Phoenix liquid handling system (Rigaku-MSC) using a ratio of 1:1 for protein/reservoir and final volumes of 0.2 and 0.4 μl for every condition. Plates were incubated at room temperature in a Gallery DT plate hotel (Rigaku-MSC). *Pa*KynB crystallization occurred using the reservoir condition 0.1 M Hepes, pH 7.5, 20 % (w/v) PEG 4000 and 10 % (v/v) propan-2-ol. Crystals grew to a maximum dimension of approximately 0.25 mm, over 5 days. The *Bc*KynB crystal was obtained using a reservoir comprising 10–16 % (w/v) PEG 3350, 5 mM $CoCl_2$, 5 mM $CdCl_2$, 5 mM $MgCl_2$ and 5 mM $NiCl_2$. Before data collection, the crystals were soaked for a few seconds in the reservoir liquid with added 25 % (v/v) glycerol and then flash frozen in liquid nitrogen. Diffraction data for *Pa*KynB and *Bc*KynB were collected using the Diamond Light Source beamline IO3 using a Pilatus 6M-F detector, indexed and integrated using XDS [13], scaled and analysed with SCALA and POINTLESS [14] from the CCP4 suite [15].

The initial crystallization conditions for *Ba*KynB gave only small mechanically twinned samples and optimization was carried out in 24-well hanging drop plates using a ratio of 1:1 in final volumes of 2 and 4 μl at 20 °C. *Ba*KynB was incubated with 8 mM L-kynurenine for 10 min at room temperature, and a structure was obtained from a crystal grown in a drop equilibrating with a reservoir containing 100 mM Tris/HCl, pH 8.5, 150 mM $MgCl_2$, 30 % (w/v) PEG 4000 and 1.5 % (v/v) dioxane. The presence of dioxane was crucial to improve the morphology such that single-crystal blocks could be obtained. Crystals of

*Ba*KynB were flash frozen directly from the drops in which they grew. Diffraction data were obtained using the in-house X-ray facility (Rigaku M007HF X-ray generator with a Saturn 944HG+ CCD detector). Data were indexed and integrated using iMOSFLM [16] and scaled and analysed by AIMLESS from the CCP4 suite [15]. To obtain crystals of the *Ba*KynB–2-aminoacetophenone complex, protein was incubated with 5% (v/v) 2-aminoacetophenone before the crystallization plate was set up. The ligand was previously dissolved to a final concentration of 20% (v/v) into buffer A2 containing 20% (w/v) PEG 4000. The reservoir contained 140 mM $MgCl_2$, 30% (w/v) PEG 4000 and 100 mM Tris/HCl, pH 8.5. Diffraction data for the *Ba*KynB–2-aminoacetophene complex were obtained at the Diamond Light Source using beamline IO3.

The structure of *Pa*KynB was determined by molecular replacement (MOLREP) [17] using a single polypeptide (molecule A) for a putative metal-dependent hydrolase from *Geobacillus stearothermophilus* (PDB code 1R61). All water molecules and ions were removed from the model. The search model, which shares 25% sequence identity with *Pa*KynB, is at 2.5 Å resolution, and although Zn^{2+} is assigned in the possible active site, it is with low occupancy (0.5), with an inflated thermal parameter (*B*-factor approximately 90 Å^2) and distances to potential co-ordinating groups that are more typical of hydrogen-bonding interactions than metal ion ligand associations. There is no information concerning why Zn^{2+} was assigned here, and there are no publications on this structure. The *Bc*KynB structure was solved using a partially refined *Pa*KynB as the search model (64% sequence identity), and then the *Ba*KynB structure was solved using *Bc*KynB (40% sequence identity) as the search model. Similar protocols were applied in the refinement of all structures. This involved first a rigid body refinement as part of the molecular replacement calculations, then iterative cycles of restrained refinement combining REFMAC5 [18] with electron and difference density map inspections, and model manipulations in COOT [19]. The starting *B*-factors for each model were derived from the Wilson *B*-factor. There are multiple polypeptide chains in the asymmetric units (two subunits for *Pa*KynB and four subunits for *Bc*KynB, *Ba*KynB and the *Ba*KynB–2-aminoacetophenone complex), and tight non-crystallographic symmetry (NCS) restraints were imposed at the onset of refinement, which were gradually released during the process. Although L-kynurenine was present in the crystallization mixture used for *Ba*KynB, there was no evidence for the enzyme product in electron density maps. Even with a much greater amount of 2-aminoacetophenone present, we only observed ordered binding of ligand in one active site. Once the protein models were complete, alternative side-chain conformers, water molecules, metal ions and ligands were incorporated into the models. A correction for a twinning component of 0.15, calculated in REFMAC, was applied in refinement of the *Ba*KynB–2-aminoacetophenone complex. Refinements were terminated when there were no significant changes in R_{work} and R_{free} values and inspection of the difference density maps suggested that no further corrections or additions were justified.

The presence of Zn^{2+} was first suggested by anomalous difference Fourier maps (not shown) and subsequently confirmed in the *Pa*KynB and *Ba*KynB crystals by X-ray absorption near-edge structure (XANES) spectra measured at Diamond Light Source on beamline IO3 using a Vortex Silicon Drift detector and an excitation energy range from 9626.7 to 9690.14 eV (Supplementary Figure S1 at http://www.biochemj.org/bj/462/bj4620581add.htm). Since no Zn^{2+} was added during protein production, purification or crystallization, then it is most likely to be derived from the media used to culture the *E. coli*.

The presence of Cd^{2+} and Zn^{2+} in the active site of *Bc*KynB was confirmed on the basis of the electron density and anomalous scattering. Positive density in a difference Fourier map calculated for the 12 sulfur atoms in the asymmetric unit gave 1.1 σ per electron. The average *B*-factor for these 12 atoms is 16.4 Å^2. This was used for the determination of the number of electrons present at the *Bc*KynB metal ion sites. The difference Fourier map revealed peaks with heights of 46.5 and 68.4 σ, respectively. These height values and the difference between them are constant within the four active sites in the asymmetric unit, and indicate the presence of two different metal atoms per active site. Comparing atomic numbers between the different metals present in the crystallization condition, one Cd^{2+} and one Zn^{2+} were added to these positions. The average *B*-factors refined to 11.4 and 10.1 Å^2, respectively. A *B*-factor-adjusted calculation suggests that one site is occupied by an ion consisting of approximately 28 e^- and the other about 48 e^-. Values of 28 and 46 e^- would be expected for Zn^{2+} and Cd^{2+}, respectively. An anomalous dispersion difference Fourier calculation provided a further check. At the wavelength used for diffraction measurements, $\lambda = 0.9795$ Å, the theoretical f'' values for zinc and cadmium are 2.480 and 2.132 e^- [20]. In an anomalous difference Fourier a slightly larger signal should therefore occur at the Zn^{2+}. Indeed this is observed in all four active sites of the asymmetric unit. The average heights of the peaks are 24.2 σ at Zn^{2+} and 22.5 σ at Cd^{2+} (Supplementary Figure S2 at http://www.biochemj.org/bj/462/bj4620581add.htm).

MolProbity [21] was used to assess geometry of all models. Secondary-structure determination involved use of DSSP [22] and visual inspection, and subunit interface surface area calculations were carried out with PISA [23]. Figures were prepared using ALINE [24] and PyMOL (http://www.pymol.org). The DALI server [26] was used to search the PDB for structural homologues, whereas superpositions were calculated using DALILITE [27]. Relevant crystallographic statistics and geometric details of the refined models are reported in Table 1.

Fluorescence spectroscopy with *Ba*KynB

Fluorescence measurements for *Ba*KynB were performed using a LS-55 PerkinElmer spectrometer to investigate ligand association/disassociation. To determine the peak of maximum emission due to tryptophan, the excitation wavelength used was 280 nm and the emission wavelength detection ranged from 300 to 400 nm. The sensitivity of the detector was fixed to 800 V, and the peak of maximum emission was experimentally determined at 337 nm.

The sample (2 ml) contained 20 μg of *Ba*KynB (400 nM), 0.02 M Tris/HCl, pH 7.4, buffer and 0.2 M NaCl. The two products of the KynB catalysed reaction, L-kynurenine and formate, and 2-aminoacetophenone were investigated separately. Concentrations ranged from 0 to 800 μM in the case of L-kynurenine and 2-aminoacetophenone and from 0 to 500 μM for formate. No fluorescence emission from either L-kynurenine or 2-aminoacetophenone was observed under the experimental conditions. In the presence of increasing concentrations of L-kynurenine and 2-aminoacetophenone, a decrease and shift of the maximum fluorescence intensity was observed. In the case of formate, no variation in the fluorescence intensity was detected. Data are represented as percentage of active site saturation within the different compound concentrations (Supplementary

Table 1 Crystallographic statistics

Structure	*Pa*KynB	*Bc*KynB	*Ba*KynB	*Ba*KynB–ligand
PDB code	4COB	4COG	4CO9	4CZ1
Space group	$P3_121$	$P2_1$	$P2_1$	$P2_1$
Wavelength (Å)	0.9795	0.9795	1.5418	0.9791
Unit cell dimensions *a*, *b*, *c* (Å)	112.7, 112.7, 90.76	76.86, 50.12, 35.2, $\beta = 94.14°$	73.17, 66.02, 83.76, $\beta = 90.32°$	73.69, 66.56, 84.06, $\beta = 90.24°$
Resolution range* (Å)	28.90–2.37	28.37–1.60	42.48–1.95	42.64–2.25
Number of reflections	133546	479924	202165	104599
Unique reflections	27365	134904	58336	38046
Completeness (%)	99.2 (94.5)	99.3 (99.8)	99.4 (94.1)	98.1 (97.7)
R_{merge}†	0.057 (0.571)	0.068 (0.477)	0.063 (0.166)	0.164 (0.551)
Redundancy	4.9	3.6	3.5	2.7
$<I/\sigma(I)>$	16.8 (2.4)	12.8 (3)	11.1 (4.3)	4.8 (2.2)
Wilson *B* (Å^2)	43.87	14.97	11.84	15.33
R_{work}‡/R_{free}§	0.1519/0.1945	0.1489/0.1842	0.1715/0.2066	0.1837/0.2241
Number of residues/waters/ligands and metals	412/181/6	831/1068/38	829/906/18	825/793/1/11
Diffraction precision indicator‖ (Å)	0.198	0.068	0.137	0.086
Bond lengths (Å)/angles¶ (°)	0.018/1.890	0.025/2.429	0.008/1.322	0.017/1.684
Average *B*-factors (Å^2)	51.4	16.4	15.3	20.9
Protein atoms	3207	6533	6658	6521
Water molecules	181	1068	906	411
Metal ions	4 Zn^{2+}	4 Zn^{2+}, 4 Cd^{2+}, 8 Mg^{2+}	8 Zn^{2+}, 5 Mg^{2+}	8 Zn^{2+}, 3 Mg^{2+}
Ligands	12 Glycerol, 10 1,2-ethanediol, 1 PEG	2 Dioxane, 3 1,2-ethanediol	2-Aminoacetophenone	
Ramachandran analyses				
Favoured regions (%)	96.1	96.1	97.2	96.5
Allowed regions (%)	100	100	100	100

*Values in parentheses refer to the highest resolution shell.
†$R_{merge} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $<I(hkl)>$ is the mean value of $I_i(hkl)$ for all *i* measurements.
‡$R_{work} = \Sigma_{hkl}||F_o| - |F_c||/\Sigma|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor.
§R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from the refinement calculations.
‖Diffraction Precision Index [42].
¶[43].

Table 2 Catalytic parameters

V_{max}, maximum catalytic velocity; K_m, Michaelis-Menten constant; specific activity (SA), amount of active enzyme from the total protein quantity; k_{cat}, number of substrate molecules that are transformed per active site and per time unit. k_{cat}/K_m defines the catalytic efficiency.

Enzyme	V_{max} (nmol · min^{-1})	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} · s^{-1}) × 10^4	SA (μM · min^{-1} · mg^{-1})
*Ba*KynB	65.41 ± 2.64	0.40 ± 0.05	50.56	12.64	130.82
*Bc*KynB	58.15 ± 0.98	0.57 ± 0.02	43.94	7.71	116.30
*Pa*KynB-His	147.99 ± 8.42	0.98 ± 0.13	114.21	11.65	295.98

Figures S3 and S4 at http://www.biochemj.org/bj/462/bj4620581add.htm), and disassociation constants were determined.

RESULTS AND DISCUSSION

Enzyme purification and quaternary structure

Highly efficient recombinant protein expression systems for the enzymes from *B. anthracis* (*Ba*KynB), *B. cenocepacia* (*Bc*KynB) and *P. aeruginosa* (*Pa*KynB) were prepared. The total protein yields, after purification, from the bacterial cultures were about 10 mg·l^{-1} for *Pa*KynB and *Bc*KynB, and 50 mg·l^{-1} for *Ba*KynB. The theoretical polypeptide mass of KynB is about 23 kDa. During the final chromatography step of purification, i.e. SEC, all three proteins were eluted as a single species with approximate mass of 40 kDa and native gels also identified that a single dimeric species was observed in each case.

Enzyme assays

The kinetic properties of the enzymes were determined using a spectrophotometric assay and are reported in Table 2. The assays confirmed that active *bone fide* KFases had been purified; for example, in the case of *Ba*KynB, approximate values for K_m 0.4 μM, V_{max} 65 nmol· min^{-1}, k_{cat} 50 s^{-1} and specific activity 130 μM min^{-1}· mg^{-1} were derived. No substrate inhibition was observed under the assay conditions and NFK concentrations that were used. In the case of *Pa*KynB, assays were carried out using protein still carrying the His-tag due to aggregation problems when removal of the tag was attempted. The kinetic parameters for *Pa*KynB, *Ba*KynB and *Bc*KynB are similar to each other. Comparisons of kinetic properties of enzymes is often complicated by the use of different assay conditions and this applies to studies with KynB from *Streptomyces parvulus* [11] and *Ralstonia metallidurans* [1] which were conducted at 30 °C and 37 °C, respectively. In these cases, the activities are

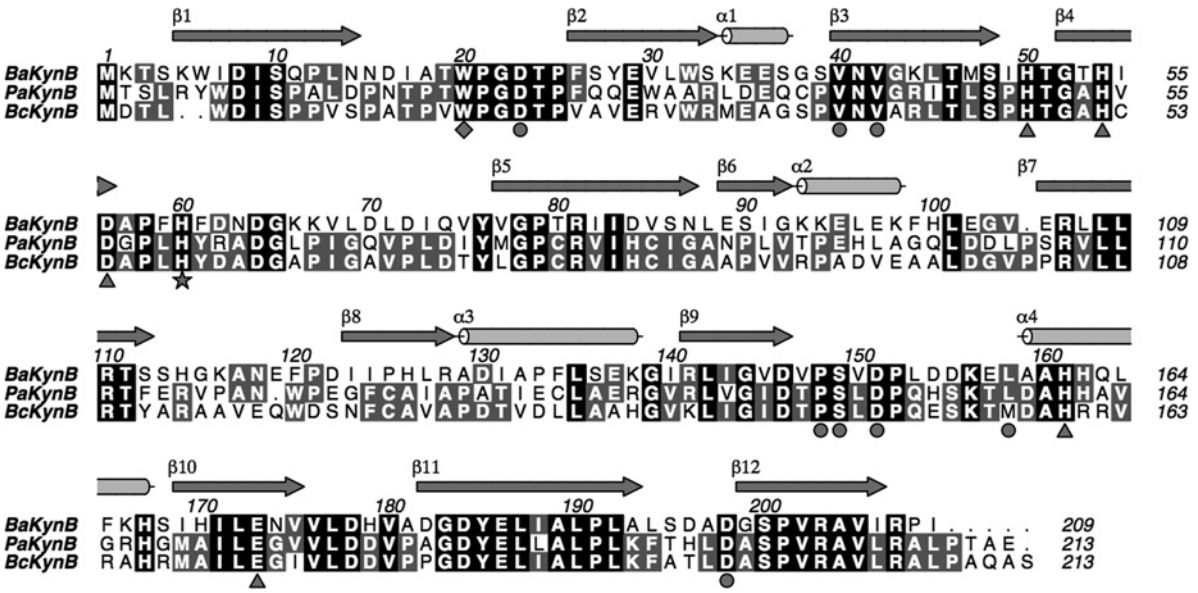


Figure 1 Sequence alignments and assignment of secondary structure for *Ba*KynB

Protein sequence alignment from *Ba*KynB, *Pa*KynB and *Bc*KynB. Blue arrows (β-sheets) and red cylinders (α-helixes) indicate the secondary structure of *Ba*KynB. Triangles mark the metal-binding amino acids. Circles mark amino acids in the active site. The star marks a key amino acid for the reaction, His^{60}. The diamond marks the tryptophan located at the active site pocket.

lower than we report with reduced specific activities of about 27 and 4 units$\cdot mg^{-1}$ noted. A more similar room-temperature assay was reported for KynB from *Bacillus cereus* [12] and this gave a specific activity of 68.5 μM $min^{-1}\cdot mg^{-1}$ with k_{cat}/K_m of 18.3×10^4 $M^{-1}\cdot s^{-1}$, values comparable with those from our observations.

Comments on crystallographic analyses and metal ion identification

The crystal structures of all three enzymes were determined (Table 1). The first analysis, *Pa*KynB at about 2.4 Å resolution identified two cations in the active site, tentatively assigned as a binuclear Zn^{2+} site. A more accurate model was sought, and the structure of *Bc*KynB was determined at 1.6 Å resolution. However, these highly ordered crystals could only by obtained in the presence of different transition metal cations including Co^{2+}, Ni^{2+} and Cd^{2+}. In addition, glycerol, the optimized cryo-protectant, was bound to the metal ions, blocking the active site, so preventing ligand studies. A structure was required without Cd^{2+} being present, or indeed any transition metal ions other than Zn^{2+} and for which we could obtain data to inform on substrate recognition. Highly ordered structures of *Ba*KynB and a complex with 2-aminoacetophenone at resolutions of 1.9 and 2.3 Å resulted and confirmed the binuclear Zn^{2+} environment. XANES scans confirmed the presence of Zn^{2+} in crystals of *Pa*KynB and in *Ba*KynB (Supplementary Figure S1).

The overall structure of KynB

The secondary and subunit structure of KynB, NCS of the dimer and detailed architecture of the active site are all highly conserved for the three enzymes. The average NCS values for overlay of Cα atoms range from 0.22 Å, for *Bc*KynB, to 0.32 Å for both *Ba*KynB and *Pa*KynB. The Gram-negative KynB structures are more similar to each other (rmsd overlay of Cα atoms is 0.70 Å) than to the Gram-positive *Ba*KynB (rmsd 1.25 Å in each case). This maps to sequence identities of about 64 % for the enzymes from the Gram-negative organisms to 40 % when compared with Gram-positive KynB (Figure 1). The structures are, however, so similar, particularly the detail in the active site which is highly conserved (see below), that we primarily detail *Ba*KynB mainly because for that example we also have data that inform on the molecular recognition of ligands in the active site.

The KynB subunit, approximate mass 23 kDa, comprises just over 200 residues with about 50 % in 12 β-strands and 15 % in four α-helices (Figures 1 and 2). Three short segments of 3_{10}-helix occur between β4 and β5. The molecule gives the appearance of a distorted eight-stranded β-barrel by treating β1 and β4 as a single strand. Three α-helices cluster on one side of the β-barrel, and on the other side is the dimer interface. The fold is not common and a search of the PDB reveals a significant similarity to only three other proteins for which there are little or no published biochemical data. These orthologues are a putative metal-dependent hydrolase/cyclase (PDB codes 3KRV and 1R61, *Z*-scores 27, 26, sequence identity 25 %), isatin hydrolase carrying a single Mn^{2+} (PDB code 4J0N, *Z*-score 23, sequence identity 23 %) and a hypothetical protein with no metal ion present (PDB code 2B0A, *Z*-score 20, sequence identity 23 %).

The KynB dimer (Figure 2) displays comparatively high values, between 22 % and 27 %, for the accessible surface area of subunits that interact with the partner. This is consistent with the observation of only dimeric species in SEC and native gels. The dimer is stabilized by hydrophobic interactions primarily involving side chains from strands β1, β11 and β12, together with a contribution from short helical segments between residues 68 and 74. The formation of an antiparallel four-stranded β-sheet involving β2–β3 from each subunit is a pronounced feature of the dimer. The N-terminal segment of β3 contributes to formation of the active site, which is mainly created by the partner subunit. In the dimer, the two catalytic sites are separated by about 30 Å.

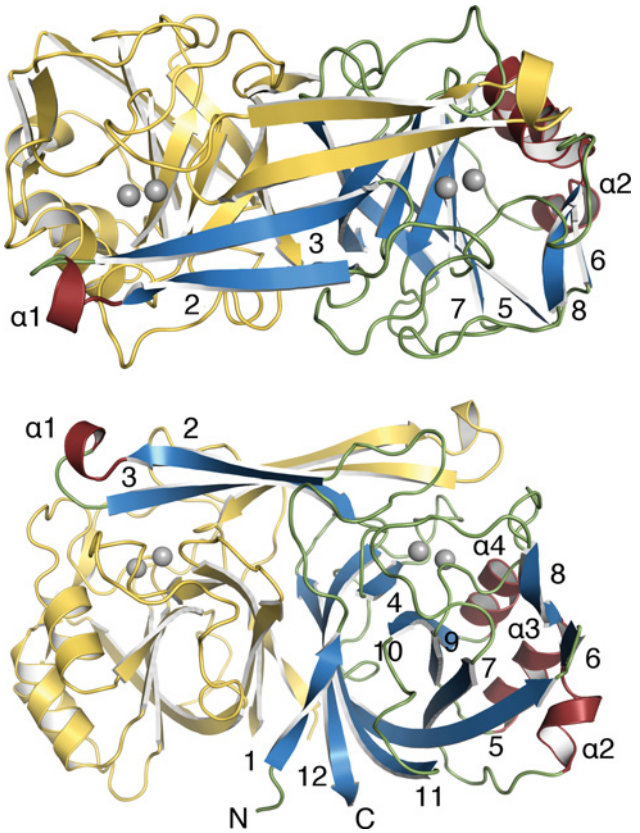


Figure 2 Ribbon diagram of the *Ba*KynB and location of the active site, orthogonal views

Zn^{2+} ions are shown as grey spheres. For one subunit, the terminal positions of the polypeptide are labelled N and C, helices are labelled and β-strands numbered. (**A**) Top view, (**B**) side view.

The active site of KynB

The enzyme active site is a narrow cavity approximately 7 Å × 12 Å and 10 Å in depth, lined by hydrophobic walls composed of Trp^{20}, His^{60} and Leu^{158} from one subunit, and Val^{40} and Val^{42} from the partner (Figure 3). Four of these residues are strictly conserved in *Bc*KynB and *Pa*KynB, whereas $Leu1^{58}$ corresponds to Met^{157} in *Bc*KynB (Figure 1). At the base of the cavity is a polar floor, where two Zn^{2+} ions, separated by a distance of 3.1 Å, are bound. This separation is unusually close for a binuclear site when compared with other systems [28].

The co-ordination environment at the binuclear Zn^{2+} site is highly organized (Figure 3). One Zn^{2+} is co-ordinated by Asp^{56}, His^{161}, Glu^{173} and a water molecule/hydroxide ion in a tetrahedral fashion with distances less than 2.3 Å. The water/hydroxide co-ordinates with the other cation and acts as a bridge for the binuclear site. A water molecule and the other oxygen of the side chain of Glu^{173} are about 2.5 Å from the cation, and together with the adjacent metal ion crowd around this Zn^{2+} to provide a distorted octahedral environment. In a similar fashion, the other cation also has four ligands co-ordinating with distances less than 2.3 Å. These are His^{50}, His^{54}, Asp^{56} and the bridging water/hydroxide. Then, another water and OE1 of Glu^{173} are 2.4 Å distant from this Zn^{2+}. The water molecules that co-ordinate the metal ions also interact by hydrogen-bond formation with the enzyme. This involves interactions with the main chain carbonyl of Pro^{148} and NE1 of Trp^{20} in one case and the side chain of Ser^{149} in the other. The bridging water or hydroxide forms a hydrogen bond with NE2 His^{60}. Such interactions may help to stabilize the active site configuration. In similar fashion, the three co-ordinating histidine side chains are held in place by hydrogen bonds that link Asp^{151}, Asp^{198} with His^{161} and His^{50}, respectively, then His54 interacts with the carbonyl group of Gly^{52}. All of these residues and their contributions to the organization of the active site are highly conserved in the three types of KynB structures characterized (Figures 1 and 4).

There are 380 unique sequences annotated as a bacterial KFase in UniProt [29]. The list is reduced to 336 entries when filtered for those with $\geqslant$30 % sequence identity with *Ba*KynB using the NCBI BLAST server [30]. The sequences were aligned (Clustal Omega) [31,32] and the conservation of key active site residues investigated. His^{54}, Asp^{56} and His^{60} are strictly conserved in all sequences, whereas His^{50}, His^{161} and Glu^{173} are maintained in 99.7 % of the entries. Trp^{20}, which contributes a significant hydrophobic character to the substrate-binding site, is strictly conserved in 81 % of the entries. Conservative substitutions for phenylalanine and tyrosine account for a further 17.8 % of the entries, confirming the importance of an aromatic group at this position.

Tryptophan fluorescence

Having identified Trp^{20} in the active site, we used fluorescence spectroscopy to investigate binding of L-kynurenine and formate, the products of the formamidase reaction. We also investigated 2-aminoacetophenone given chemical similarities to part of L-kynurenine. This ligand provides the aromatic component of substrate and product and primarily lacks only the 2-amino-4-oxobutanoic acid moiety. Formate did not elicit any spectroscopic change but L-kynurenine and 2-aminoacetophenone gave comparable K_d values of approximately 60 and 50 μM, respectively (Supplementary Figures S3 and S4). A *Ba*KynB–2-aminoacetophenone complex structure then revealed aspects of molecular recognition in the active site (Figure 3 and Supplementary Figure S5 at http://www.biochemj.org/bj/462/bj4620581add.htm). As we will show, this is consistent with the ligands binding in the vicinity of the active site tryptophan with comparable affinity and suggests that 2-aminoacetophenone represents a suitable molecule from which we can derive information on aspects of substrate recognition. Formic acid did not elicit any change in tryptophan fluorescence, suggesting that this moiety of substrate and one product of the KynB-catalysed reaction does not interact with the active site tryptophan.

Recognition of a conformationally restricted substrate is implied

We were unable to obtain the structure of a complex of KynB with the products of catalysis, formate or kynurenine. However, the characterization of a *Ba*KynB–2-aminoacetophenone complex structure at 2.25 Å resolution provides a suitable template to inform on aspects of molecular recognition in the active site (Figure 3). Key to ligand, and by implication substrate, binding is the interaction with Trp^{20}. There are also van der Waals associations with Val^{40} from the partner subunit. The 2-amino group forms a hydrogen bond to the Zn^{2+} bridging water or hydroxide. In this respect, the complex derived may closely mimic the configuration at the completion of the catalytic reaction and immediately prior to the replacement of product by the incoming substrate for another round of catalysis. An intramolecular hydrogen bond between the amine and carbonyl groups of 2-aminoacetophenone appeared to be significant. A 3D search of the Cambridge Structural Database [33] using 2-aminoacetophenone

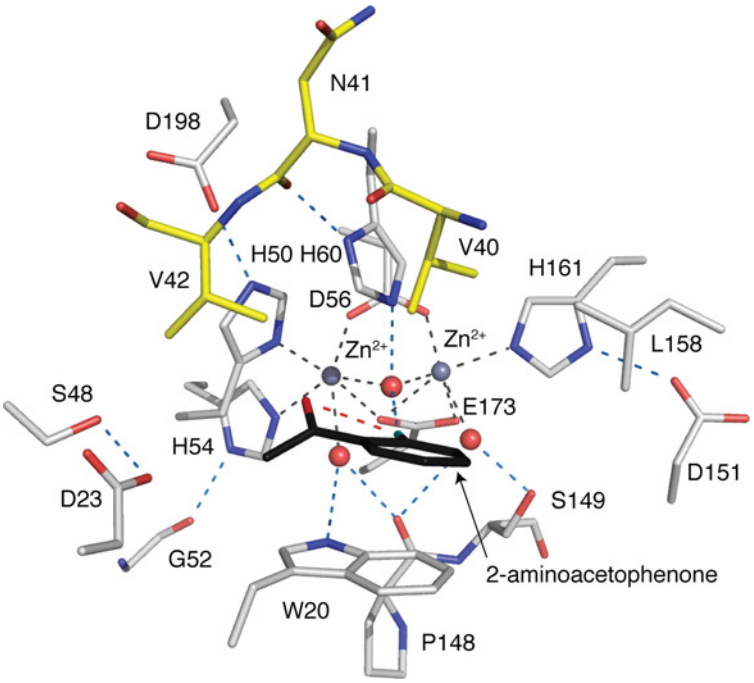


Figure 3 Active site of the *Ba*KynB–2-aminoacetophenone complex

Zn^{2+} is a grey sphere, grey broken lines mark co-ordination to amino acid side chains and waters/hydroxide (red spheres). Amino acid atomic positions are coloured with C in grey or yellow depending on subunit, N in blue, O in red, and 2-aminoacetophenone is shown with C in black, N in cyan and O in red. Blue broken lines represent potential hydrogen bonds and a single red broken line indicates the intramolecular interaction in the ligand.

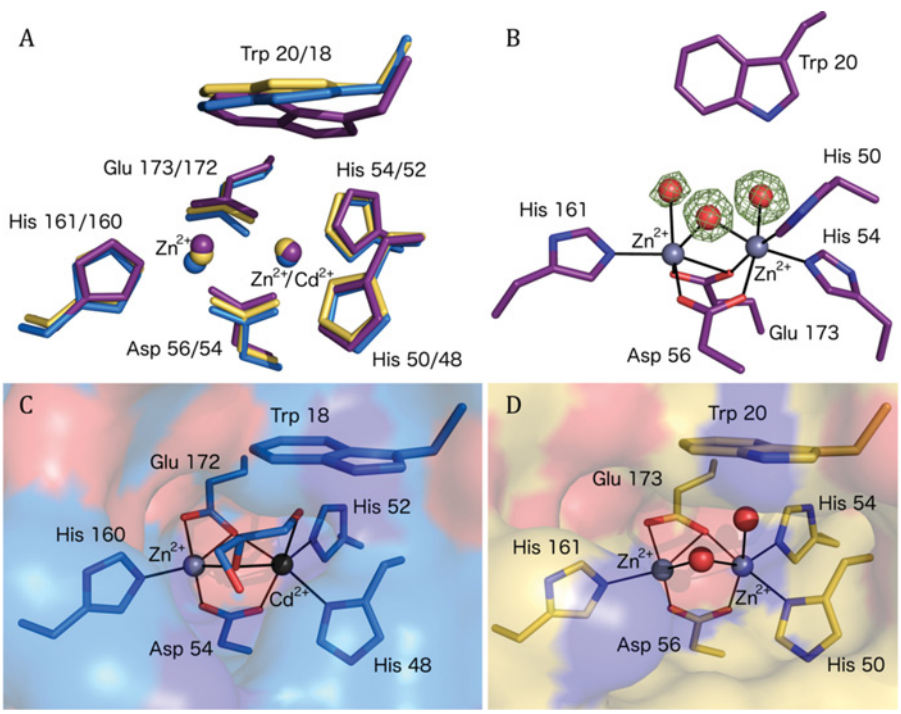


Figure 4 The active site of KynB is highly conserved

(**A**) Superimposition of *Ba*KynB (violet), *Bc*KynB (blue) and *Pa*KynB (yellow) active site structures. (**B**) The *Ba*KynB active site with difference density for the three water molecules (red spheres) in the active site depicted as dark green chicken wire and contoured at 5σ. (**C**) The *Bc*KynB active site with glycerol bound to the metal ions. (**D**) The *Pa*KynB active site. Continuous lines represent co-ordinating contacts to the Zn^{2+} ions (grey spheres).

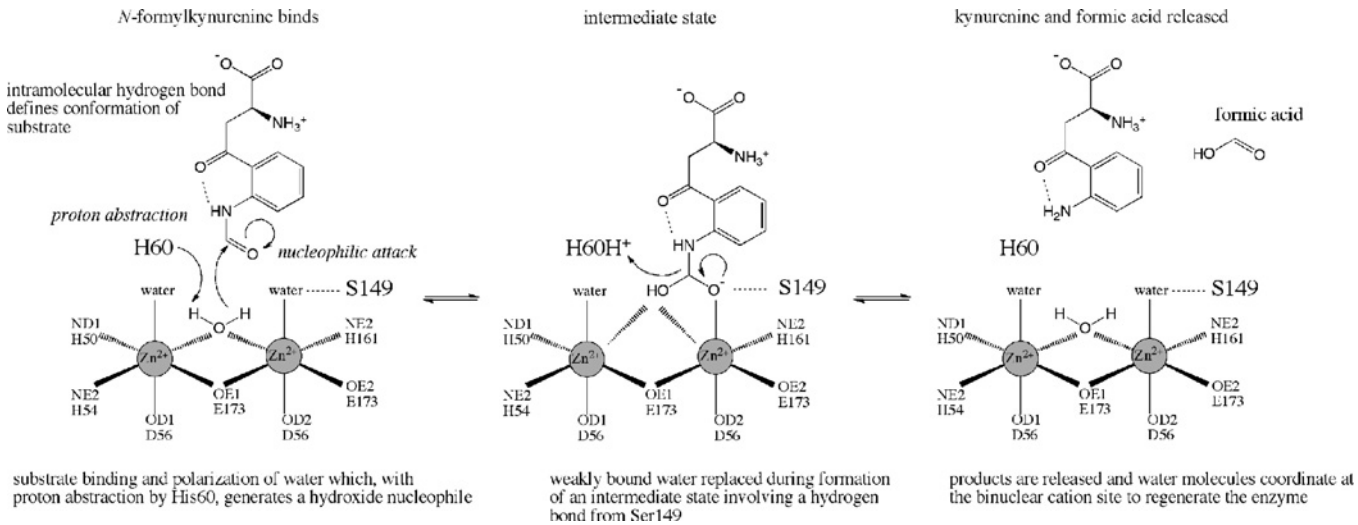


Figure 5 A proposed mechanism for the amidase KynB

(*S*)-2-Amino-4-(2-formamidophenyl)-4-oxobutanoic acid (*N*-formylkynurenine) is converted into (*S*)-2-amino-4-(2-aminophenyl)-4-oxobutanoic acid (kynurenine).

as the template identified 28 similar compounds with structures determined at atomic resolution and all of which display the same interaction. Therefore, such an intramolecular hydrogen bond probably defines or constrains the substrate conformation, and orients the scissile bond such that the formyl moiety would be directed over between His[60] and Ser[149]. These are two strictly conserved residues in KynB sequences. The complex structure suggests then that when substrate binds the oxygen of the formyl substituent could displace a water that binds Zn^{2+} weakly and perhaps also interact directly with Ser[149]. The orientation or alignment is then fixed for nucleophilic attack to occur at carbon as is described further below.

Mechanistic considerations and comparisons with metallohydrolases

Although the enzyme fold is different, the binuclear Zn^{2+} catalytic centre of KynB suggests some similarity to mononuclear and binuclear metalloenzymes in the amidohydrolase superfamily. Such enzymes exploit the Lewis acid properties of the cation, decrease the pK_a of water and generate a hydroxide nucleophile to initiate catalysis [34,35]. They bind and polarize substrate during an intermediate state in the catalytic cycle and generally, then use an acidic glutamate or aspartate to donate a solvent-acquired proton as the intermediate collapses to release products. Examples of such metallohydrolases include type B β-lactamases [28,36], phosphotriesterases [37], leucine aminopeptidases [38] and dihydro-orotases [39]. Constant features of this superfamily [34] include using an acidic component as proton donor and the presence of flexible loops to engage with substrate are in particular noted in the $(\beta/\alpha)_8$ triose phosphate isomerase-type barrel structures. A relevant example actually occurs further down in the kynurenine pathway, it is the enzyme α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (ACMSD) [40,41]. ACMSD possess a single Zn^{2+} ion co-ordinated by three histidine residues and an aspartate. Other $(\beta/\alpha)_8$ barrel enzymes such as specific phosphodiesterases and dihydro-orotases also possess a carboxylated lysine, which binds two metal ions, typically Zn^{2+}, holding them about 3.5–4.0 Å apart and with enough room around the metal ions to allow for a co-ordination increase from four to five or six in the intermediate state of catalysis [34,35].

KynB immediately struck us as distinct with a binuclear Zn^{2+} site, cations only 3.1 Å apart, which is closer than in other cases, in a more crowded environment and with little scope to increase the co-ordination number given that six co-ordinating groups already surround each ion. This suggests then that straightforward replacement of a co-ordinating water ligand might be important. Furthermore, unlike other metallohydrolases where conformational flexibility appears to be important [34,36,37], the catalytic site of KynB is relatively rigid with an average thermal parameter (*B*-factor) of 9.6 Å^2 compared with an overall average of 15.1 Å^2 in *Ba*KynB.

Based on the structural data and consistency with the fluorescence measurements, we can now suggest a plausible and distinctive mechanism (Figure 5). The fixed conformation of a planar substrate would assist binding in the small rigid active site cavity. His[60], a strictly conserved active site residue, in conjunction with the powerful Lewis acid activating binuclear Zn^{2+} environment, is ideally positioned to acquire a proton from the bridging water to generate a potent nucleophilic hydroxide for attack at the carbonyl component of substrate. Direct interaction between the substrate and Zn^{2+} as a catalytic intermediate is formed would probably involve replacement of weakly co-ordinated water assisted by interaction with Ser[149]. The intermediate would collapse as His[60], which in the complex is 3.6 Å from the 2-amino group of the ligand, donates the proton to form the amine. The C–N bond breaks and products are released.

AUTHOR CONTRIBUTIONS

Laura Díaz-Sáez, Velupillai Srikannathasan and Martin Zoltner performed the research; Laura Díaz-Sáez, Velupillai Srikannathasan and William N. Hunter designed the research, analysed the data and wrote the paper. William N. Hunter supervised the research.

ACKNOWLEDGEMENTS

We thank Paul Fyfe, Alice Dawson and Thomas Eadsforth for advice, and the Diamond Light Source Synchrotron Radiation Facility for beam time.

FUNDING

This research was supported by the Wellcome Trust [grant numbers 082596, 094090 and 100476] and a studentship from the Defence Science and Technology Laboratory (U.K.).

REFERENCES

1 Kurnasov, O., Jablonski, L., Polanuyer, B., Dorrestein, P., Begley, T. and Osterman, A. (2003) Aerobic tryptophan degradation pathway in bacteria: novel kynurenine formamidase. FEMS Microbiol. Lett. **227**, 219–227 CrossRef PubMed
2 Zummo, F. P., Marineo, S., Pace, A., Civiletti, F., Giardina, A. and Puglia, A. M. (2012) Tryptophan catabolism via kynurenine production in *Streptomyces coelicolor*: identification of three genes coding for the enzymes of tryptophan to anthranilate pathway. Appl. Microbiol. Biotechnol. **94**, 719–728 CrossRef PubMed
3 Stone, T. W., Stoy, N. and Darlington, L. G. (2013) An expanding range of targets for kynurenine metabolites of tryptophan. Trends Pharmacol. Sci. **34**, 136–143 CrossRef PubMed
4 Kurnasov, O., Goral, V., Colabroy, K., Gerdes, S., Anantha, S., Osterman, A. and Begley, T. P. (2003) NAD Biosynthesis: identification of the tryptophan to quinolinate pathway in bacteria. Chem. Biol. **10**, 1195–1204 CrossRef PubMed
5 Wogulis, M., Chew, E. R., Donohoue, P. D. and Wilson, D. K. (2008) Identification of formyl kynurenine formamidase and kynurenine aminotransferase from *Saccharomyces cerevisiae* using crystallographic, bioinformatic and biochemical evidence. Biochemistry **47**, 1608–1621 CrossRef PubMed
6 Han, Q., Robinson, H. and Li, J. (2012) Biochemical identification and crystal structure of kynurenine formamidase from *Drosophila melanogaster*. Biochem. J. **446**, 253–260 CrossRef PubMed
7 Fyfe, P. K., Oza, S. L., Fairlamb, A. H. and Hunter, W. N. (2008) *Leishmania* trypanothione synthetase-amidase structure reveals a basis for regulation of conflicting synthetic and hydrolytic activities. J. Biol. Chem. **283**, 17672– 17680 CrossRef PubMed
8 Fyfe, P. K., Rao, V. A., Zemla, A., Cameron, S. and Hunter, W. N. (2009) Specificity and mechanism of *Acinetobacter baumanii* nicotinamidase: implications for activation of the front-line tuberculosis drug pyrazinamide. Angew. Chem. Int. Ed. **48**, 9176–9179 CrossRef PubMed
9 Pace, H. C. and Brenner, C. (2001) The nitrilase superfamily: classification, structure and function. Genome Biol. **2**, 1–9 CrossRef PubMed
10 Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy serve. In The Proteomics Protocols Handbook (Walker, J. M., ed.), pp. 571–607, Humana Press, Totowa CrossRef
11 Katz, E., Brown, D. and Hitchcock, M. J. (1987) Arylformamidase from *Streptomyces parvulus*. Methods Enzymol. **142**, 225–234 CrossRef PubMed
12 Bougie, J. M. (2011) Expression, Purification, and Characterization of Kynurenine Formamidase from *Bacillus cereus*. Masters Thesis, George Manson University, Fairfax, VA, U.S.A.
13 Kabsch, W. (2010) XDS. Acta Crystallogr. D Biol. Crystallogr. **66**, 125–132 CrossRef PubMed
14 Evans, P. R. (2006) Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. **62**, 72–82 CrossRef PubMed
15 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A. et al. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. **67**, 235–242 CrossRef PubMed
16 Leslie, A. G. W. and Powell, H. R. (2007) In Evolving Methods for Macromolecular Crystallography (Read, R. J. and Sussman, J. L., eds), pp. 41–51, Springer, Dordrecht CrossRef
17 Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. J. Appl. Crystallogr. **30**, 1022–1025 CrossRef
18 Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. **67**, 355–367 CrossRef PubMed
19 Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. **66**, 486–501 CrossRef PubMed
20 Sasaki, S. (1989) Numerical Tables of Anomalous Scattering Factors Calculated by the Cromer and Liberman's Method (KEK Report 88–14). National Laboratory for High Energy Physics, Tsukuba
21 Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. **66**, 12–21 CrossRef PubMed
22 Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22**, 2577–2637 CrossRef PubMed
23 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. **372**, 774–797 CrossRef PubMed
24 Bond, C. S. and Schüttelkopf, A. W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr. D Biol. Crystallogr. **65**, 510–512 CrossRef PubMed
25 Reference deleted
26 Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res. **38**, W545–W549 CrossRef PubMed
27 Hasegawa, H. and Holm, L. (2009) Advances and pitfalls of protein structural alignment. Curr. Opin. Struct. Biol. **19**, 341–348 CrossRef PubMed
28 Schenk, G., Mitić, N., Gahan, L. R., Ollis, D. L., McGeary, R. P. and Guddat, L. W. (2012) Binuclear metallohydrolases: complex mechanistic strategies for a simple chemical reaction. Acc. Chem. Res. **45**, 1593–1904 CrossRef PubMed
29 The UniProt Consortium (2014) Activities at the universal protein resource (UniProt). Nucleic Acids Res. **42**, 191–198 CrossRef PubMed
30 Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T. L. (2008) NCBI BLAST: a better web interface. Nucleic Acids Res. **36**, W5–W9 CrossRef PubMed
31 Sievers, F., Wilm, A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. **7**, 539 CrossRef PubMed
32 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) Jalview version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics **25**, 1189–1191 CrossRef PubMed
33 Thomas, I. R., Bruno, I. J., Cole, J. C., Macrae, C. F., Pidcock, E. and Wood, P. A. (2010) WebCSD: the online portal to the Cambridge Structural Database. J. Appl. Crystallogr. **43**, 362–366 CrossRef PubMed
34 Holm, L. and Sander, C. (1997) An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. Proteins **28**, 72–82 CrossRef PubMed
35 Seibert, C. M. and Raushel, F. M. (2005) Structural and catalytic diversity within the amidohydrolase superfamily. Biochemistry **44**, 6383–6392 CrossRef PubMed
36 Bebrone, C. (2007) Metallo-β-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. Biochem. Pharmacol. **74**, 1686–1701 CrossRef PubMed
37 Bigley, A. N. and Raushel, F. M. (2013) Catalytic mechanisms for phosphotriesterases. Biochim. Biophys. Acta **1834**, 443–453 CrossRef PubMed
38 Sträter, N., Sun, L., Kantrowitz, E. R. and Lipscomb, W. N. (1999) A bicarbonate ion as a general base in the mechanism of peptide hydrolysis by dizinc leucine aminopeptidase. Proc. Natl. Acad. Sci. U.S.A. **96**, 11151–11156 CrossRef PubMed
39 Lee, M., Maher, M. L., Christopherson, R. I. and Guss, J. M. (2007) Kinetic and structural analysis of mutant *Escherichia coli* dihydroorotases: a flexible loop stabilizes the transition state. Biochemistry **46**, 10538–10550 CrossRef PubMed
40 Huo, L., Fielding, A. J., Chen, Y., Li, T., Iwaki, H., Hosler, J. P., Chen, L., Hasegawa, Y., Que, L. Jr and Liu, A. (2012) Evidence for a dual role of an active site histidine in alpha-amino-beta-carboxymuconate-epsilon-semialdehyde decarboxylase. Biochemistry **24**, 5811–5821 CrossRef PubMed
41 Garavaglia, S., Perozzi, S., Galeazzi, L., Raffaelli, N. and Rizzi, M. (2009) The crystal structure of human alpha-amino-beta-carboxymuconate-epsilon-semialdehyde decarboxylase in complex with 1,3-dihydroxyacetonephosphate suggests a regulatory link between NAD synthesis and glycolysis. FEBS J. **276**, 6615–6623 CrossRef PubMed
42 Cruickshank, D. W. J. (1999) Remarks about protein structure precision. Acta Crystallogr. D Biol. Crystallogr. **55**, 583–601 CrossRef PubMed
43 Engh, R. A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr. A **47**, 392–400 CrossRef

Received 17 April 2014/27 May 2014; accepted 19 June 2014
Published as BJ Immediate Publication 19 June 2014, doi:10.1042/BJ20140511

Biochem. J. (2014) **462**, 581–589 (Printed in Great Britain) doi:10.1042/BJ20140511

online data

SUPPLEMENTARY ONLINE DATA

Structures of bacterial kynurenine formamidase reveal a crowded binuclear zinc catalytic site primed to generate a potent nucleophile

Laura DÍAZ-SÁEZ*, Velupillai SRIKANNATHASAN*, Martin ZOLTNER* and William N. HUNTER*[1]

*Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, U.K.

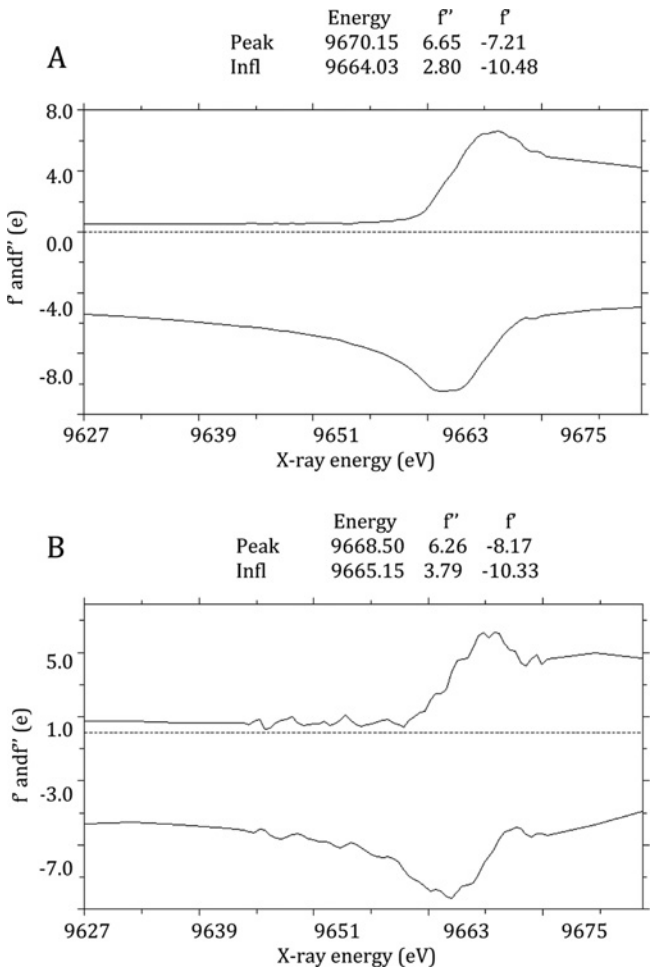


Figure S1 XANES spectra

K-edge XANES spectra from *Ba*KynB (**A**) and *Pa*KynB (**B**) showing the anomalous scattering factor *f*′ and *f*″ values within a range of X-ray energy (from 9626.7 to 9690.14 eV).

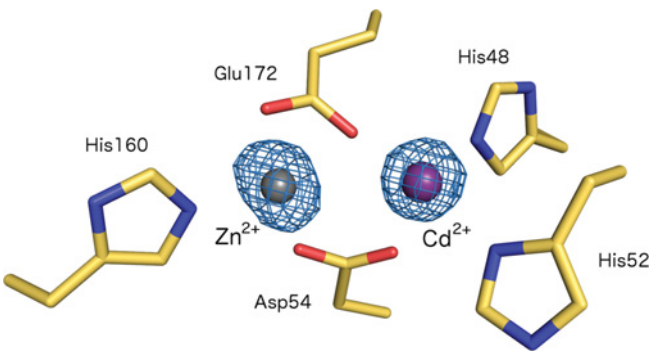


Figure S2 Anomalous difference Fourier map for *Bc*KynB

The map, blue chicken wire, is contoured at 6σ. Zn^{2+} and Cd^{2+} are shown as grey and purple spheres respectively. Water molecules and glycerol have been omitted for the purpose of clarity.

[1] To whom correspondence should be addressed (email w.n.hunter@dundee.ac.uk).

Atomic co-ordinates and structure factors have been deposited in the PDB under codes 4COG, 4COB, 4CO9 and 4CZ1.

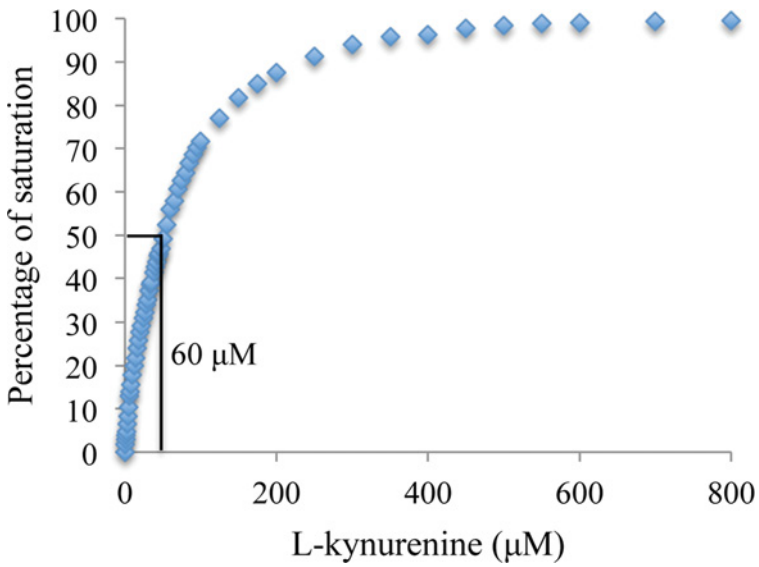


Figure S3 L-Kynurenine binding

Plot derived from fluorescence spectroscopy showing the percentage of active site saturation at different L-kynurenine concentrations.

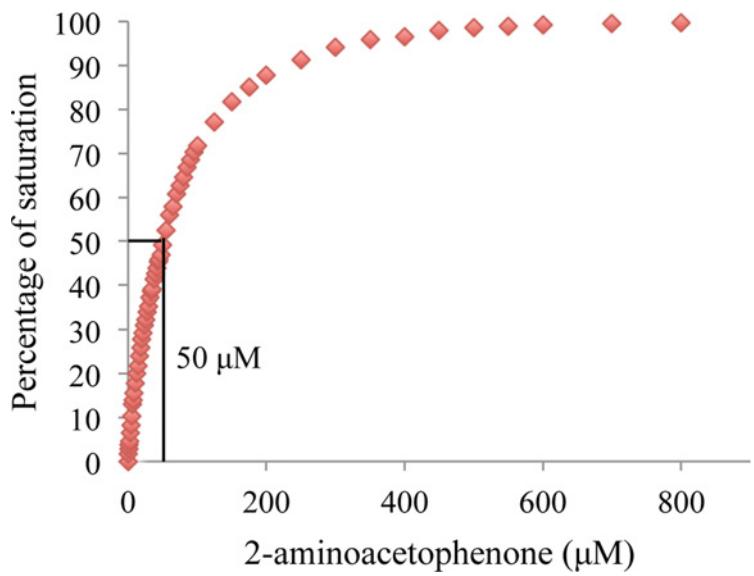


Figure S4 2-Aminoacetophenone binding

Plot derived from fluorescence spectroscopy showing the percentage of active site saturation at different 2-aminoacetophenone concentrations.

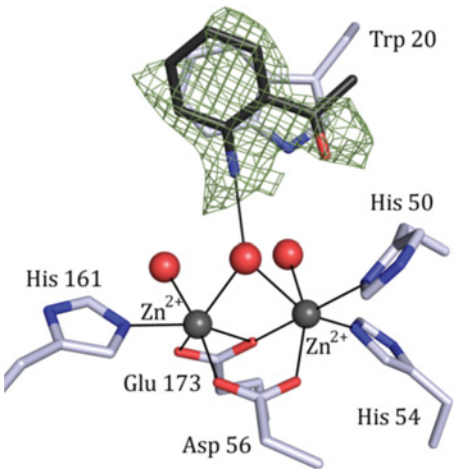


Figure S5 Omit difference density for 2-aminoacetophenone

The $F_o - F_c$ omit map is shown as green chicken wire and contoured at 1.5σ.

Received 17 April 2014/27 May 2014; accepted 19 June 2014
Published as BJ Immediate Publication 19 June 2014, doi:10.1042/BJ20140511